

A unified framework for harmonic analysis of functions on directed graphs and changing data

H. N. Mhaskar*

Abstract

We present a general framework for studying harmonic analysis of functions in the settings of various emerging problems in the theory of diffusion geometry. The starting point of the now classical diffusion geometry approach is the construction of a kernel whose discretization leads to an undirected graph structure on an unstructured data set. We study the question of constructing such kernels for directed graph structures, and argue that our construction is essentially the only way to do so using discretizations of kernels. We then use our previous theory to develop harmonic analysis based on the singular value decomposition of the resulting non-self-adjoint operators associated with the directed graph. Next, we consider the question of how functions defined on one space evolves to another space in the paradigm of changing data sets recently introduced by Coifman and Hirn. While the approach of Coifman and Hirn require that the points on one space should be in a known one-to-one correspondence with the points on the other, our approach allows the identification of only a subset of landmark points. We introduce a new definition of distance between points on two spaces, construct localized kernels based on the two spaces and certain interaction parameters, and study the evolution of smoothness of a function on one space to its lifting to the other space via the landmarks. We develop novel mathematical tools that enable us to study these seemingly different problems in a unified manner.

Keywords: Kernel construction, directed graphs, changing data problems, extension problems, wavelet-like representation, Tauberian theorem.

1 Introduction

There are many approaches developed during the last decade or so in order to analyze large, unstructured, possibly high dimensional data. The basic idea is to embed the high dimensional data on a low dimensional sub-manifold of the ambient Euclidean space. The main theme of the research is then to understand the data geometry in terms of the geometric properties of this sub-manifold. Well known techniques in this direction, dimensionality reduction in particular, are Isomaps [63], maximum variance unfolding (MVU) (also called semidefinite programming (SDP)) [66], locally linear embedding (LLE) [57], local tangent space alignment method (LTSA) [67], Laplacian eigenmaps (Leigs) [3], Hessian locally linear embedding (HLLE) [19], diffusion maps (Dmaps) [16], and randomized anisotropic transform [11]. A recent survey of these methods is given by Chui and Wang in [12]. An excellent introduction to the subject of diffusion geometry can be found in the special issue [8] of Applied and Computational Harmonic Analysis, 2006. The application areas are too numerous to mention exhaustively. They include, for example, document analysis [17], face recognition [38, 46, 11], hyperspectral imaging [10], semi-supervised learning [2, 4], image processing [22, 6], cataloguing of galaxies [23], and social networking [64].

The starting point in diffusion geometry is the point cloud. A point cloud is a set \mathcal{P} of points $\{x_i\}$ in a Euclidean space, together with a similarity relation W , viewed as the matrix of edge weights in an undirected graph. In the absence of any known structure on the set of points, the similarity relation is constructed by the so-called diffusion matrix. For example, a standard construction for W is given by

$$W_{i,j} = \exp(-\|x_i - x_j\|^2/\epsilon),$$

where $\{x_i\}$ is the set of points, $\|\cdot\|$ is the Euclidean norm, and ϵ is a judiciously chosen variance parameter. The first few eigenvalues and eigenvectors of this matrix (or some related matrix, such as $\text{diag}(\text{row sums of } W) - W$)

*Department of Mathematics, California Institute of Technology, Pasadena, CA 91125; Institute of Mathematical Sciences, Claremont Graduate University, Claremont, CA 91711. The research of this author is supported in part by ARO Grant W911NF-15-1-0385. email: hrushikesh.mhaskar@cgu.edu.

provide the desired low dimensional embedding. Usually, these vectors codify certain identifiable features of the data set. For example, in [11], the point cloud consists of thumbnail images of the faces of the same person in different orientations. The components of the first non-trivial eigenvector orders the collection according to the angle of rotation. Many such examples can be found in the literature; indeed, the well known concept of kernel PCA is based on this fact. It is proved in [3, 5, 45, 61] that as the data becomes dense, the so-called graph Laplacian based on a judiciously chosen (weighted) adjacency matrix, such as W , converges to the Laplace–Beltrami operator on the **unknown** manifold from which the data is sampled, and likewise for the corresponding eigenvalues/eigenfunctions. A deeper insight into the intimate connection between the eigenfunctions and the manifold is explained in [40, 41], where it is shown that some of the eigenfunctions define a local coordinate system on the unknown manifold so that the Euclidean distance between the points represented with these coordinates is proportional to the geodesic distance between the points.

The applications of this theory to semi-supervised learning can be formulated as problems of function extension, going beyond the important and difficult question of understanding the data geometry. For example, in semi-supervised learning for classification, the class labels are known only on a small *training subset* $\mathcal{C} \subset \mathcal{P}$, and the objective is to find the class labels for all points in \mathcal{P} . The class label for a point x can be viewed as the value of a *target function* f at x , and the question is then to extend f from \mathcal{C} to \mathcal{P} . Formulated in this manner, one can even think of extending f to the entire manifold \mathbb{X} , including those points which are not in the original data set \mathcal{P} . Clearly, all the regression problems in learning theory are also problems of function extension/approximation, for example [30]. Indeed, one of the main reasons for the popularity of neural and radial basis function networks in learning theory is their universal approximation property. An important problem in this paradigm is to estimate the fidelity of the approximation scheme at unseen data points. Within the context of diffusion geometry, a rigorous analytic theory to address this question is developed in [48, 25, 26, 51, 50], based on an abstract framework formulated in [52]. The constructive algorithm that emerges from this theory is tested in the proof-of-concept experiments on recognition of hand-written digits [48], the Cleveland heart disease data set, the Wisconsin breast cancer data set, and a new experiment at NIH to predict automatically the local classifications of drusen in patients with age related macular degeneration (AMD) based on multi-spectral fundus images of the retina [24].

The goal of the present paper is to develop a general framework in which function extension/approximation problems can be studied in the context of two of the newly emerging paradigms in the theory of diffusion geometry, namely, the problem of changing data sets, and the problem of analyzing functions on a directed graph.

The problem of analyzing changing data arises in many applications, for example, modeling of social networks, where the relationships between the people involved may change over time, analysis of financial markets, evolutionary biological questions, and the analysis of medical tests in patients as they change over a time period in the life of the patient. In classical diffusion geometry, the eigenfunctions and eigenvalues are based on an existing data set. The out-of-sample extension problem is the problem of incorporating new data as they emerge. This problem can be viewed clearly as the problem of changing data as well.

Our technical motivation for the analysis of changing data sets is the paper [15] of Coifman and Hirn. In this paper, the authors define a diffusion distance between points lying on two different manifolds, as well as a distance between changing manifolds, based on these. One important feature of this framework is that the points in the point clouds (respectively, the underlying manifolds) remain the same over different manifolds, and can be identified uniquely in a one-to-one correspondence as the manifold changes. The authors illustrate an application of this framework in the context of a sequence of hyperspectral images of essentially the same area over time. Each image is treated as a point cloud, where the “points” are the spectral curves. Each point in one image is identified with exactly one point in another image at the same spatial pixel location.

In practical applications, it is often not possible to identify each point on one manifold uniquely with the corresponding point on another manifold. For example, in the analysis of brain MRI images of Alzheimer patients taken over time, the images are taken at random points of the brain [42]. Therefore, even if the brain surfaces are the “same” in the sense of belonging to the same patient, it is not possible to identify one point in the image uniquely with another point on another image. In fact, there are different number of points in different images. One may, however, consider the cortical thickness of the brain as a function on an imaginary sphere, approximate its values at certain standard points, such as dyadic points, which can then be identified from one sphere to the other in the sequence of images as they evolve in time.

In a different example, if one considers each frame of a video to be a point cloud/manifold as is often done in image processing applications of diffusion geometry, each point on one video frame can get mapped to different points on the next frame with different probabilities [53]. However, it may be reasonable to assume that certain “landmark” points can be identified from manifold to manifold as they change. A related direction of research is manifold matching, where one needs to match features of one manifold with the other to determine whether the

two manifolds arise from different representations of the same data.

The framework developed in this paper enables us to study the lifting of functions from one manifold to another via a set of landmarks rather than requiring a complete one-to-one correspondence between points of the different manifolds.

Another important emerging direction is to extend the diffusion geometry paradigm to directed graphs rather than undirected graphs defined by a symmetric matrix such as W above. For example, graphs arising from many applications such as transportation problems in which traffic flow is restricted to one direction must be directed graphs, in order to allow analysis of traffic along one way streets. In analyzing data about world-wide web links, the direction of the links contains crucial information. Similarly, in an understanding of blood circulation, the direction of blood flow is important. Other applications of directed graphs include one-way communication problems, social interactions between people where the relationship is not necessarily symmetric, and scheduling problems.

Directed graphs are, of course, studied in great detail over more than a century. In particular, the notion of symmetric graph Laplacian for a directed graph whose weighted adjancancy matrix is given by a Markov matrix is developed by F. Chung in [13]. A recent well written Ph. d. thesis by Gidelew [29] develops harmonic analysis based on this Laplacian, and cites several other papers in this direction. While this line of research assumes that edge weights in the graph are already known, it is a separate question to impose a directed graph structure for an unstructured data, analogous to the graph structure implied on undirected data by matrices such as W . One may view the asymmetric affinity matrices in [14] as kernels defining a directed graph in the abstract sense. To the best of our knowledge, the only deliberate effort in this direction is the recent paper [54] by Mousazadeh and Cohen, where a modified Gaussian kernel is proposed, and an analogue of the limiting Laplace–Beltrami operator is presented. The theory of non-self-adjoint integral and differential operators is quite well developed, e.g., [31, 36, 47]. It is therefore natural to obtain the desired graph structure by discretizing a suitable kernel, deemed most appropriate for the application in question.

Unlike [3, 5, 45, 61] in the case of undirected graphs or [54] in the case of directed graphs, it is not our intention to focus on specific constructions of the kernels which give rise to some concrete differential operator in the limiting sense. The book [47] of Ma and Röckner gives a very readable account of the interconnection between non-self-adjoint differential operators, the associated Dirichlet forms, Green’s kernels, and semi-groups of operators generated by the differential operators. Several examples are also discussed in [47], and may be adapted for this purpose. Instead, we wish to enumerate certain properties that such kernels ought to have in order to develop a reasonably rich theory of harmonic analysis of functions defined on changing data or directed graphs. We will, however, make some comments on what we consider likely to be a good construction of kernels in these cases, leaving the details in each case to future research. In particular, in the case of construction of a directed graph structure on data, we will argue that our suggestions provide essentially the only way to have such a structure, as well as enable the construction of such structure to allow meaningful harmonic analysis based on this structure.

To describe our results more precisely, it is convenient to review the setting in the “classical” diffusion geometry theory [17]. Let \mathbb{X} be a metric measure space with a metric d and a sigma-finite Borel measure μ^* . Let $\{\phi_k\}$ be a complete orthonormal system in $L^2(\mu^*)$, considered as the eigenfunctions of a Laplacian operator defined by a formally defined kernel

$$-\sum_{k=0}^{\infty} \lambda_k^2 \phi_k(x) \overline{\phi_k(y)}, \quad x, y \in \mathbb{X},$$

where λ_k is a sequence of non-negative numbers with $\lambda_k \uparrow \infty$ as $k \rightarrow \infty$. For the analysis of a measurable function $f : \mathbb{X} \rightarrow \mathbb{C}$, Coifman and Maggioni [17] propose taking dyadic powers of the “diffusion” operator

$$f \mapsto \int_{\mathbb{X}} f(y) \left\{ \sum_{k=0}^{\infty} \exp(-\lambda_k^2 t) \phi_k(\cdot) \overline{\phi_k(y)} \right\} d\mu^*(y), \quad x \in \mathbb{X},$$

for a small value of t . In practice this is done as dyadic powers of a sparse matrix using fast algorithms. Theoretically, the range of the 2^{-2n} -th power of this operator is approximately equal to the space

$$\Pi_{2^n \epsilon} = \text{span}\{\phi_k : \lambda_k < 2^n \epsilon\}.$$

for a suitable ϵ . The corresponding projection operator is given by

$$f \mapsto \int_{\mathbb{X}} f(y) \left\{ \sum_{k: \lambda_k < 2^n \epsilon} \phi_k(\cdot) \overline{\phi_k(y)} \right\} d\mu^*(y).$$

For the detection of certain anomalies in the target function, it is desirable to approximate the function in the uniform norm rather than the $L^2(\mu^*)$ norm. It is well known that the projection operators are usually unbounded if uniform approximation is desired, even in the very classical case of trigonometric Fourier series. In [52, 48], we have initiated a study of function approximation and harmonic analysis in $L^p(\mu^*)$ for values of $p \neq 2$; with a special emphasis on uniform approximation. The important starting point of our theory is therefore to construct localized operators of the form

$$f \mapsto \int_{\mathbb{X}} f(y) \left\{ \sum_{k=0}^{\infty} h\left(\frac{\lambda_k}{n}\right) \phi_k(\cdot) \overline{\phi_k(y)} \right\} d\mu^*(y) \quad (1.1)$$

for a suitable (low pass or band pass) “filter” h .

Taking a sufficiently smooth function h as the filter ensures that these operators are uniformly bounded in every $L^p(\mu^*)$, $1 \leq p \leq \infty$. Using these operators, one can obtain wavelet-like dyadic representations in all of these spaces, where the behavior of the terms characterizes completely the smoothness properties of the target function f . One fascinating feature of this construction is that even if the operator itself is defined using the spectral information about the target function f , the localization estimate leads one to the analogues of the well known theorems of Jaffard [18, Theorems 9.2.1, 9.2.2]; i.e., the local smoothness of the target function at a point can be characterized completely in terms of the behavior of the terms of the wavelet-like representation in a neighborhood of the point in question. It is proved in [9] that such localized harmonic analysis can always be done when appropriate localized kernels are available.

A feature common to both the new directions of research mentioned above that distinguishes the new problems from the classical one is that one has to deal with two systems jointly. Thus, in the case of changing data or manifold matching, one has two admissible systems $\Xi_1 = (\mathbb{X}_1, d_1, \mu_1^*, \{\lambda_{1,k}\}, \{\phi_{1,k}\})$ and $\Xi_2 = (\mathbb{X}_2, d_2, \mu_2^*, \{\lambda_{2,k}\}, \{\phi_{2,k}\})$ (cf. Definition 2.1 for a precise definition). The objective of interest in [15] was to define a “diffusion distance” between points on the two manifolds using the respective heat kernels on the two manifolds. In our paper, the goal is to study how a function defined on one system evolves into another function on the other system via an interaction between the various parameters defining the two systems. We will assume the interaction only in an abstract sense, without requiring that each point on one manifold should be identified uniquely with another point on the other manifold. Moreover, the construction and theory is general enough to deal with situations where a point on one manifold may go to different points on the other manifold according to some probability distribution.

The case of directed graphs is much easier in some sense. A directed graph is generated in this paradigm by the discretization of a non-self-adjoint kernel operator. Thus, we again have two admissible systems, where the underlying measure spaces, measures, and distances are the same in both the systems. In analogy to the classical diffusion setting, it is convenient to take this kernel to be a suitable kernel related to some differential operator on the underlying space, and the sequence $\{\lambda_k\}$ to be the singular values of the differential operator. The two systems arise by taking for $\{\phi_{1,k}\}$, $\{\phi_{2,k}\}$ the left and right singular functions of the operator. With this insight, there is essentially only one way to construct such operators, stemming from the polar decomposition of the non-self-adjoint operator involved, and the theory which we have already developed can be applied directly with minor changes.

The basic objective of this paper is to obtain the analogues of the localized operators analogous to those in (1.1) for harmonic analysis in these settings. Obviously, it cannot hold that these operators applied to a function on say \mathbb{X}_2 , will converge to the function itself – the result of this application is not even a function on \mathbb{X}_2 . However, under certain conditions, one can obtain a lifting operator from one space to the other so as to result in a lifting of the original starting function to a corresponding function on the other space, and the smoothness properties of this lifted function can be described using those of the starting function. A motivating model here is the classical theorems in the theory of partial differential equations (or integral equations), such as trace theorems, inverse theorems, or embedding theorems [56].

The important technical tool for achieving the unified framework is to prove what is known as a Tauberian theorem. To describe this idea, we make an observation in the classical setting. If one considers the univariate function defined by

$$\mu(u) = \mu(x, y; u) = \sum_{k: \lambda_k < u} \phi_k(x) \overline{\phi_k(y)}, \quad u > 0,$$

then μ has bounded total variation on compact sub-intervals of $[0, \infty)$, and the projection kernel can be expressed formally in the form of the Stieltjes integral

$$\sum_{k: \lambda_k < n} \phi_k(x) \overline{\phi_k(y)} = \int_0^\infty h(u/n) d\mu(u),$$

where $h(t) = 1$ if $0 \leq t < 1$, and $= 0$ otherwise. The kernel of the operators defined in (1.1) can also be described in the same way with different choices of h . Likewise, the heat kernel can be expressed formally as a Stieltjes integral

$$\int_0^\infty \exp(-u^2 t) d\mu(u).$$

Our theorems in [48, 25] on the localization properties of the kernel of the operators defined in (1.1) can therefore be viewed as statements about the properties of one integral transform of μ in terms of those of another transform. Such a theorem is called a Tauberian theorem; a famous example being the Wiener–Ikehara Tauberian theorem regarding the density of translates of a fixed function in $L^1(\mathbb{R})$. The history of such theorems in the settings of classical harmonic analysis spans over a century, and is summarized excellently in a recent book [44] by Korevaar. Early versions of the Tauberian theorems of the form which we will prove in this paper were obtained by Bochner and Chandrasekharn [7].

Viewed in this abstraction, the specific nature of μ becomes irrelevant, in particular, the dependence of μ on x, y or systems $\{\lambda_k\}$, $\{\phi_k\}$ does not play any role anymore. Therefore, the problem of constructing localized kernels in two or more systems reduces to the problem of proving an appropriate Tauberian theorem. While the classical development of such theorems is focused on the limiting or asymptotic behavior of a measure μ itself near ∞ or some summability operator applied to μ , our theorem has a very different character. We are interested in obtaining some very specific bounds on one transform when some other very specific bounds on the other transform are known. In particular, the proof relies heavily upon both complex variable theory as well as distribution theory.

We summarize the main new contributions of this paper as follows:

- We point out a very simple way to construct kernels to embed a directed graph into a manifold, and argue that this is essentially the **only way** to construct such kernels.
- We develop a theory for harmonic analysis of functions on directed graphs, whether using existing kernels as the starting point, or by constructing specially designed kernels.
- We develop a theory of lifting a function on one data space to another using only the identification of certain interaction parameters connecting the two spaces.
- We develop a mathematical theory that enables us to treat both of these seemingly unrelated problems in a unified manner. Other contemplated applications of this theory include an analysis of solutions of partial differential equations with non-self-adjoint operators, bit representation of signals on Euclidean spaces, and smoothness-preserving image completion.

In Section 2, we will present our suggestions for the construction of kernels whose discretizations yield the desired directed graph structure on an unstructured data. We will argue that our construction is essentially the only one based on the idea of discretizing operators, barring some esoteric counterexamples. We will then discuss harmonic analysis on the systems generated by these non-self-adjoint kernels. The results are partly a review of our previous results, and partly a bridge between these and the results on changing data, which will be described in Section 3. The proofs of all the results, including the Tauberian theorem will be given in Section 4.

2 Directed graphs

In this section, we focus on a discussion of harmonic analysis on directed graphs. In Sub-section 2.1, we discuss the construction of the systems themselves, either from a pre-existing structure, or from the point of view of constructing kernels in order to embed the given data as a directed graph on points on an ambient manifold. We will point out that there is essentially only one way to make such constructions. In Sub-section 2.2, we assume that the system is constructed, and study the harmonic analysis of functions on the graphs based on the system assuming certain conditions. The results are obtained as immediate consequences of our results in [48, 25], and serve both as a review of these results and a prelude to the results to be presented in Section 3. We note that in the continuous setting, it is natural to think of a directed graph as a non-self-adjoint operator. Therefore, it is not necessary that this operator must map a Hilbert space to itself. For example, in applications in photo-acoustic tomography one encounters operators defined on one Hilbert space, taking values in a different Hilbert space [27]. We will explain in Remark 3.2 how the analysis presented in this section can be extended easily to such situations. However, for simplicity, we will restrict ourselves in this section to the case of one space.

2.1 Construction of the graph structure

To motivate our discussion, let us recall that a directed graph with the vertex set $\{x_i\}_{i=1}^M$, and edge weights $w(x_i, x_j)$ is an $M \times M$ non-symmetric matrix W whose entries are $w(x_i, x_j)$. Then the polar form for W [39, Corollary 7.3.3] is PU , where P is a uniquely determined positive semi-definite matrix with the same rank as W , and U is unitary and real. If W has full rank, then U is determined uniquely. In this case, efficient algorithms to compute the polar decomposition are given recently by Nakatsukasa, Bai, and Gygi [55].

Writing P in the form

$$P(x_i, x_j) = \sum_{k=0}^{M-1} \lambda_k \phi_k(x_i) \phi_k(x_j),$$

for some orthogonal vectors ϕ_k , the singular value decomposition of W is given by

$$W(x_i, x_j) = \sum_{k=0}^{M-1} \lambda_k \phi_k(x_i) (U^* \phi_k)(x_j).$$

Thus, if f is any function on the points $\{x_i\}$, then

$$(Wf)(x_i) = (P(Uf))(x_i).$$

Therefore, the harmonic analysis of f based on the singular vectors of W is the same as the harmonic analysis of Uf based on the eigenvectors of P , as in [52, 37]. Moreover, this is essentially the only shape such a harmonic analysis can take.

We digress to make a comparison with the construction of the graph Laplacian for directed graphs proposed in [13], where the role of W is played by a non-symmetric matrix denoted in [13] by $\Phi^{1/2} P \Phi^{-1/2}$. Writing in this discussion, I for the identity matrix, the graph Laplacian in [13] is then given by

$$I - \frac{W + W^*}{2} = 2 \left(I - \left(\frac{W + I}{2} \right) \left(\frac{(W + I)^*}{2} \right) \right) - \frac{1}{2}(I - WW^*),$$

which is the same for the graph Laplacian corresponding to the underlying undirected graph given by $(W + W^*)/2$. The quantity $P^2 = WW^*$ in our notation refers to a different undirected graph. If $W_{i,j}$ denotes the transition probability of going from state i to state j , $(WW^*)_{i,j}$ denotes the probability of nodes i and j leading to some common state. The matrix $(1/2)(W + I)((1/2)(W + I))^*$ can also be given a similar interpretation, where self-loops are added to the original graph at each vertex.

In the case of infinite graphs, the same analysis and logic works, provided the infinite matrix of edge weights represents a compact operator. The operator U in this case may only be a partial isometry rather than a unitary operator, but this detail does not affect our analysis in this section. In this case, we can view the matrix as the kernel, and the underlying measure to be an atomic measure [36]. If we do not have a choice of the weight matrix, then clearly, the problem of finding the polar decomposition is equivalent to the problem of finding the singular value decomposition. However, the basic fact that the resulting harmonic analysis will be the same as the harmonic analysis of Uf based on the symmetric component is still the same.

In the case when there is no preconceived notion of what the weight matrix ought to be, and we have to impose the structure as in the case of classical diffusion geometry, these observations make the problem of kernel construction almost trivial. We start with any kernel based, self-adjoint, positive semi-definite operator, with any familiar kernel used in classical diffusion geometry that leads to a compact operator, such as the Gaussian, or some Green's kernel of a desired elliptic differential operator, take a composition with an appropriate partial isometry (even a unitary operator for simplicity), and discretize the resulting operator. This approach leads to a very flexible construction, where also the unitary operator may be treated as a parameter to be chosen according to the application. In particular, if P is an operator that commutes with a self-adjoint operator L then it is obvious that W and LU commute in the sense that $(LU)^*W = W^*(LU)$ and $W(LU)^* = (LU)W^*$.

We observe that this is essentially the **only** way to construct a reasonable directed graph structure in the paradigm of diffusion geometry; i.e., by discretizing kernel based operators. For example, if the non-self-adjoint operator is a Hilbert-Schmidt operator, then a polar decomposition is guaranteed [31], and the fact that the self-adjoint operator is also Hilbert-Schmidt, and hence, kernel based, is observed, for example, in [36].

Thus, a simple example of the construction in [54] is the following. We wish to construct an operator on a Euclidean space \mathbb{R}^q . Fixing $z^* \in \mathbb{R}^q$ and $t > 0$, we define

$$(W_t f)(x) = (2\pi t)^{-1/2} \int_{\mathbb{R}^q} \exp\left(-\frac{\|x - y - z^*\|^2}{t}\right) f(y) dy = (2\pi t)^{-1/2} \int_{\mathbb{R}^q} \exp\left(-\frac{\|x - y\|^2}{t}\right) f(y - z^*) dy.$$

It is readily seen that this has the polar decomposition as above where the positive definite self-adjoint operator is defined by the Gaussian kernel, and the unitary operator is defined by $f \mapsto f(\cdot - z^*)$. Greater flexibility is obtained by treating z^* as a parameter in addition to t . In [54], the integral is taken over an unknown manifold, the inner product $z^* \cdot (x - y)$ is replaced by a vector field operating on $x - y$, and some numerical experiments are presented to illustrate the idea.

2.2 Harmonic analysis

In this section, we will assume that the necessary structures are available, and develop harmonic analysis based on these. This section may be viewed partly as a review of our previous work on one system, and partly as an example of the results in Section 3. For this reason, some of the definitions will be more general than necessary in the context of directed graphs.

First, we describe some terminology.

If \mathbb{X} is a measure space and ν is a (signed complex valued, or positive) measure on \mathbb{X} , then for measurable $f : \mathbb{X} \rightarrow \mathbb{C}$, we define

$$\|f\|_{\mathbb{X}, \nu, p} = \begin{cases} \left\{ \int_{\mathbb{X}} |f(x)|^p d|\nu|(x) \right\}^{1/p}, & \text{if } 1 \leq p < \infty, \\ |\nu| - \text{ess sup}_{x \in \mathbb{X}} |f(x)|, & \text{if } p = \infty. \end{cases} \quad (2.1)$$

The space $L^p(\mathbb{X}, \nu)$ consists of all (equivalence classes of) functions $f : \mathbb{X} \rightarrow \mathbb{C}$ for which $\|f\|_{\mathbb{X}, \nu, p} < \infty$. If the space \mathbb{X} is clear from the context, then we will drop its mention from the notation, and similarly if the measure ν is clear from the context, we will drop its mention as well, as long as the mention of \mathbb{X} is dropped. Thus, for example, $L^p(\mu^*)$ refers to $L^p(\mathbb{X}, \mu^*)$.

We recall that if \mathbb{X} is a measure space with a positive measure μ^* , a system $\{\phi_k\}_{k=0}^\infty \subset L^2(\mu^*)$ is called a **Bessel system** if there is a subset \mathcal{D} of $L^1(\mu^*) \cap L^\infty(\mu^*)$, dense in $L^2(\mu^*)$, and a functional \mathcal{N} on $L^2(\mu^*)$ such that

$$\sum_{k=0}^\infty \left| \int_{\mathbb{X}} f(x) \overline{\phi_k(x)} d\mu^*(x) \right|^2 \leq \mathcal{N}(f), \quad f \in \mathcal{D}. \quad (2.2)$$

For example, if \mathbb{X} is a complete Riemannian manifold without boundary, and μ^* is its volume measure, F is a vector field defined on \mathbb{X} , and $\{\phi_k\}$ are the eigenfunctions of the Laplace-Beltrami operator on \mathbb{X} then the Green's formula implies that $\{F\phi_k\}$ is a Bessel system.

Definition 2.1 *Let \mathbb{X} be a locally compact quasi-metric space with a sigma-finite positive Borel measure μ^* , d be the quasi-metric on \mathbb{X} , $\{\lambda_k\}$ be a non-decreasing sequence of positive numbers, with $\lambda_0 = 0$ and $\lim_{k \rightarrow \infty} \lambda_k = \infty$. Let $\{\phi_k\}$ be a Bessel system of functions in $L^2(\mathbb{X}, \mu^*)$, such that each ϕ_k is continuous, bounded, and integrable on \mathbb{X} . Then $\Xi = (\mathbb{X}, d, \mu^*, \{\lambda_k\}, \{\phi_k\})$ will be called an **admissible system**.*

In this section, we assume that a directed graph is represented in abstract by an admissible system $\Xi = (\mathbb{X}, d, \mu^*, \{\lambda_k\}, \{\phi_k\})$ (cf. Definition 2.1), where $\{\phi_k\}$ is an orthonormal set of functions in $L^2(\mu^*)$, together with a partial isometry operator U on $L^2(\mu^*)$. The measure μ^* may be a discrete measure, and we do not assume that the system is constructed with any particular choice of kernels. Therefore, our theorems are applicable in the settings of finite or infinite pre-defined directed graphs or in the setting where a directed graph structure is constructed in order to embed the graph on a manifold, as long as the required assumptions to be enumerated in Theorem 2.1 are satisfied. The system $\{\psi_k = U^* \phi_k\}$ is necessarily also an orthonormal system, so that the system $\Xi' = (\mathbb{X}, d, \mu^*, \{\lambda_k\}, \{\psi_k\})$ is also an admissible system. As we pointed out in Section 2.1, the harmonic analysis of a function f on \mathbb{X} in the setting of this abstract directed graph is actually a decomposition of Uf into a convergent series of components band limited to different frequency bands with respect to the system $\{\phi_k\}$, based on the data on f with respect to the system $\{\psi_k\}$. We will show that the behavior of these components characterize the smoothness properties of Uf .

In the absence of any differentiability structure, one needs to define the notion of smoothness carefully. We will do this for the system Ξ , but will need to use it also for the system Ξ' . First, we define for $f \in L^1(\mu^*) + L^\infty(\mu^*)$,

$$\hat{f}(k) = \hat{f}(\Xi; k) = \int_{\mathbb{X}} f(y) \overline{\phi_k(y)} d\mu^*(y), \quad k = 0, 1, \dots \quad (2.3)$$

The analogue of the classical Sobolev classes with smoothness index γ is the class of all functions in $L^2(\mu^*)$ for which

$$\sum_{k=0}^\infty (k^2 + 1)^\gamma |\hat{f}(k)|^2 < \infty.$$

This definition has the advantage that it does not depend upon any differentiability structure on \mathbb{X} . However, even in the most classical case of the trigonometric Fourier analysis, it is well known that the Fourier coefficients do not characterize the smoothness of the function in spaces other than the space of 2π -periodic functions square integrable on $[-\pi, \pi]$. So, this definition does not quite work in the case of other L^p spaces. To modify this definition, we will use the notion of the degree of approximation, $E_{n,p}(\Xi; f)$ defined in (2.5) below.

For $n > 0$ (not necessarily integer), let

$$\Pi_n = \Pi_n(\Xi) = \text{span}\{\phi_k : \lambda_k < n\}, \quad (2.4)$$

$$E_{n,p}(f) = E_{n,p}(\Xi; f) = \inf_{P \in \Pi_n} \|f - P\|_{\mathbb{X}, \mu^*, p}, \quad f \in L^p(\mathbb{X}, \mu^*), \quad 1 \leq p \leq \infty. \quad (2.5)$$

The class of all f for which $E_{n,p}(f) \rightarrow 0$ as $n \rightarrow \infty$ will be denoted by $X^p = X^p(\Xi)$. In view of Parseval identity we observe that if $f \in X^2(\Xi)$, then

$$E_{n,2}^2(f) = \sum_{k: \lambda_k \geq n} |\hat{f}(k)|^2.$$

Hence, using a routine dyadic sum argument, it is easy to deduce that

$$\sum_{k=0}^{\infty} (k^2 + 1)^\gamma |\hat{f}(k)|^2 \sim \sum_{j=0}^{\infty} 2^{2j\gamma} E_{2^j,2}(f)^2.$$

(Here, and in the sequel, we use the following convention regarding constants. The symbols c, c_1, \dots will denote generic positive constants independent of the obvious variables, such as f in the above equation. The values of these constants may be different at different occurrences, even within the same formula. The symbol $A \sim B$ will mean that $c_1 A \leq B \leq c_2 B$.) This observation prompts the following definition. For $0 < \rho \leq \infty$ and $0 < \gamma < \infty$, the Besov space $B_{p,\rho,\gamma} = B_{p,\rho,\gamma}(\Xi)$ is the class of all $f \in X^p$ such that the quantity $\|f\|_{p,\rho,\gamma} < \infty$, where

$$\|f\|_{p,\rho,\gamma} = \|f\|_{\Xi,p,\rho,\gamma} = \begin{cases} \|f\|_{\mathbb{X}, \mu^*, p} + \left(\sum_{j=0}^{\infty} 2^{\gamma \rho j} E_{2^j,p}(f)^\rho \right)^{1/\rho}, & \text{if } 0 < \rho < \infty, \\ \|f\|_{\mathbb{X}, \mu^*, p} + \sup_{j \geq 0} 2^{\gamma j} E_{2^j,p}(f), & \text{if } \rho = \infty. \end{cases} \quad (2.6)$$

The index γ may be thought of as the smoothness parameter, the parameter ρ adds an additional level of refinement to the smoothness classes. It is well known in approximation theory that the smoothness of a function $f \in X^p$ as measured by K -functionals can be characterized in terms of $E_{n,p}(f)$. In the context of diffusion geometry, we have given such theorems in [48, 25].

It is convenient to define a sequence space $\mathbf{b}_{\rho,\gamma}$ consisting of all sequences $\mathbf{a} = \{a_k\}_{k=0}^{\infty}$ for which

$$\|\mathbf{a}\|_{\mathbf{b}_{\rho,\gamma}} = \begin{cases} \left(\sum_{j=0}^{\infty} 2^{\gamma \rho j} |a_j|^\rho \right)^{1/\rho}, & \text{if } 0 < \rho < \infty, \\ \sup_{j \geq 0} 2^{\gamma j} |a_j|, & \text{if } \rho = \infty, \end{cases} \quad (2.7)$$

is finite. Thus, $B_{p,\rho,\gamma}$ consists of $f \in X^p$ such that $\{E_{2^j,p}(f)\}_{j=0}^{\infty} \in \mathbf{b}_{\rho,\gamma}$.

Our analysis of functions on \mathbb{X} is based on spectral information of the form $\{\hat{f}(\Xi'; k)\}_{k=0}^{\infty}$, but the smoothness of Uf is measured in terms of the Besov spaces $B_{p,\rho,\gamma}(\Xi)$.

To obtain the wavelet-like representation, we introduce some further terminology.

For a function $H : [0, \infty) \rightarrow [0, \infty)$, we define formally a kernel

$$\Phi_n(\Xi, \Xi'; H, x, y) = \sum_{k=0}^{\infty} H\left(\frac{\lambda_k}{n}\right) \phi_k(x) \overline{\psi_k(y)}, \quad x, y \in \mathbb{X}. \quad (2.8)$$

Definition 2.2 A function $h : \mathbb{R} \rightarrow [0, 1]$ is called a **low pass filter** if h is even, $h(u) = 1$ if $|u| < 1/2$, $h(u) = 0$ if $|u| \geq 1$, and h is non-increasing on $[0, \infty)$.

With a low pass filter h , we next define a reconstruction operator for $f \in L^1(\mu^*) + L^\infty(\mu^*)$, $x \in \mathbb{X}$ by

$$\sigma_n(\Xi, \Xi'; h, f, x) = \sum_{k=0}^{\infty} h\left(\frac{\lambda_k}{n}\right) \hat{f}(\Xi'; k) \phi_k(x) = \int_{\mathbb{X}} \Phi_n(\Xi, \Xi'; h, x, y) f(y) d\mu^*(y), \quad (2.9)$$

and the analysis operators

$$\tau_j(\Xi, \Xi'; h, f, x) = \begin{cases} \sigma_1(\Xi, \Xi'; h, f, x), & \text{if } j = 0, \\ \sigma_{2^j}(\Xi, \Xi'; h, f, x) - \sigma_{2^{j-1}}(\Xi, \Xi'; h, f, x), & \text{if } j = 1, 2, \dots \end{cases} \quad (2.10)$$

We observe that since h is a low pass filter, σ_1 is a filtered projection on the low frequency components for which $\lambda_k < 1$. In the case when the λ_k 's are obtained as singular values of some operator, σ_1 clearly includes the projection on the kernel space of this operator.

In the case of analysis in $L^2(\mu^*)$, we may take h to be the usual cut-off function: $h(t) = 1$ if $0 \leq t < 1$, $h(t) = 0$ if $t \geq 1$. The operators σ_n and τ_j are then the projections on Π_n and the “wavelet space” $\Pi_{2^j} \ominus \Pi_{2^{j-1}}$ respectively. In the case of other L^p spaces, for the purpose of analysis in uniform norm in particular, we need a smooth function h . A smooth function h results in a well localized kernel for the operators.

Next, it is convenient to formulate some of our important assumptions in the form of a definition.

Definition 2.3 *The **heat kernel** on an admissible system $\Xi = (\mathbb{X}, d, \mu^*, \{\lambda_k\}, \{\phi_k\})$ is defined formally by*

$$K_t(\Xi; x, y) = \sum_{k=0}^{\infty} \exp(-\lambda_k^2 t) \phi_k(x) \overline{\phi_k(y)}. \quad (2.11)$$

We say that Ξ satisfies the **Gaussian upper bound condition** (with exponent q) if there exists $q > 0$, such that

$$|K_t(\Xi; x, y)| \leq c_1 t^{-q/2} \exp\left(-c_2 \frac{d(x, y)^2}{t}\right), \quad 0 < t \leq 1, \quad x, y \in \mathbb{X}. \quad (2.12)$$

In particular, the series in (2.11) converges for all $x, y \in \mathbb{X}$.

Example 2.1 (the manifold case) A typical example of a system satisfying the Gaussian upper bound condition is the following. Let \mathbb{X} be a compact (C^∞), connected Riemannian manifold, μ^* be the Riemannian volume measure on \mathbb{X} , and d be the geodesic distance. We consider the Laplace–Beltrami operator on \mathbb{X} , and let $\{-\lambda_k^2\}$ be the sequence of its eigenvalues, with corresponding sequence of eigenfunctions $\{\phi_k\}$. Then the system $(\mathbb{X}, d, \mu^*, \{\lambda_k\}, \{\phi_k\})$ is an admissible system. It is shown in [43] that under certain conditions on \mathbb{X} , if $-\lambda_k^2$ are the eigenvalues of certain elliptic operators, and ϕ_k are the corresponding eigenfunctions, then (2.12) is satisfied. Many other general situations in metric measure spaces are known [33, 34, 35] and references therein. \square

Remark 2.1 In the case when \mathbb{X} is a finite graph, the heat kernel corresponding to a vertex-based graph Laplacian satisfies the Gaussian upper bound condition for large values of t rather than small values as in Definition 2.3. It is shown in [28] that the solution of the wave equation corresponding to an edge-based Laplacian satisfies the property known as finite speed of wave propagation. In turn, this is equivalent to the Gaussian upper bound condition for small values of t as stipulated in Definition 2.3 [60, 25].

We will denote the ball of radius r around x by

$$\mathbb{B}(x, r) = \{y \in \mathbb{X} : d(x, y) \leq r\}, \quad (2.13)$$

and note that all balls are compact.

With this preparation, we are ready to state our wavelet-like representation theorem.

Theorem 2.1 *Let $\Xi = (\mathbb{X}, d, \mu^*, \{\lambda_k\}, \{\phi_k\})$ be an admissible system satisfying the Gaussian upper bound condition with exponent $q > 0$. Let U be a partial isometry operator on $L^2(\mu^*)$, $\psi_k = U^* \phi_k$, and $\Xi' = (\mathbb{X}, d, \mu^*, \{\lambda_k\}, \{\psi_k\})$ be the associated admissible system. We assume further that μ^* satisfies the regularity condition*

$$\mu^*(\mathbb{B}(x, r)) \leq cr^q, \quad x \in \mathbb{X}, \quad r > 0. \quad (2.14)$$

Let $1 \leq p \leq \infty$, $f \in L^1(\mu^*) + L^\infty(\mu^*)$, $Uf \in X^p$, and h be a sufficiently smooth low pass filter.

(a) For $n \geq 1$, we have

$$E_{n/2,p}(\Xi; Uf) \leq \|Uf - \sigma_n(\Xi, \Xi'; h, f)\|_{\mathbb{X}, \mu^*, p} \leq c E_{n,p}(\Xi; Uf). \quad (2.15)$$

(b) In the sense of $L^p(\mu^*)$ convergence,

$$Uf = \sum_{j=0}^{\infty} \tau_j(\Xi, \Xi'; h, f). \quad (2.16)$$

(c) If $p = 2$, then

$$\sum_{j=0}^{\infty} \|\tau_j(\Xi, \Xi'; h, f)\|_{\mathbb{X}, \mu^*, 2}^2 \leq \|Uf\|_{\mathbb{X}, \mu^*, 2}^2 \leq 5 \sum_{j=0}^{\infty} \|\tau_j(\Xi, \Xi'; h, f)\|_{\mathbb{X}, \mu^*, 2}^2. \quad (2.17)$$

(d) Let $0 < \rho \leq \infty$, $\gamma > 0$. Then $Uf \in B_{p, \rho, \gamma}$ if and only if $\{\|\tau_j(\Xi, \Xi'; h, f)\|_{\mathbb{X}, \mu^*, p}\}_{j=0}^{\infty} \in \mathbf{b}_{\rho, \gamma}$.

Remark 2.2 In the case when U is the identity operator, Theorem 2.1 reduces to the corresponding theorems in [48] for the undirected case. \square

3 Analysis on changing data

Our constructions in this section are motivated by the paper [15] of Coifman and Hirn. In our terminology, they consider two admissible systems $\Xi_1 = (\mathbb{X}, d_1, \mu^*, \{\lambda_{1,j}\}, \{\phi_{1,j}\})$ and $\Xi_2 = (\mathbb{X}, d_2, \mu^*, \{\lambda_{2,k}\}, \{\phi_{2,k}\})$ with the base measure space (\mathbb{X}, μ^*) being common to both, and $\{\phi_{1,j}\}, \{\phi_{2,k}\}$ being orthonormal systems. The diffusion distance between $x_1, x_2 \in \mathbb{X}$, with the points considered as elements of different copies of \mathbb{X} is defined by

$$\|K_t(\Xi_1; x_1, \cdot) - K_t(\Xi_2; x_2, \cdot)\|_2 = \left\{ K_{2t}(\Xi_1; x_1, x_1) + K_{2t}(\Xi_2; x_2, x_2) - 2\Re \int_{\mathbb{X}} K_t(\Xi_1; x_1, y) K_t(\Xi_2; y, x_2) d\mu^*(y) \right\}^{1/2}.$$

Thus, the interaction term is the inner product expression above, which can be written in the form

$$\Re \sum_{j,k=0}^{\infty} \exp(-t(\lambda_{1,j}^2 + \lambda_{2,k}^2)) \overline{\langle \phi_{1,j}, \phi_{2,k} \rangle} \phi_{1,j}(x_1) \overline{\phi_{2,k}(x_2)},$$

where $\langle \cdot, \cdot \rangle$ denotes in this context the inner product in $L^2(\mu^*)$.

This prompts the following more general set up. We are interested in the interaction of two systems $\Xi_1 = (\mathbb{X}_1, d_1, \mu_1^*, \{\lambda_{1,k}\}, \{\phi_{1,k}\})$ and $\Xi_2 = (\mathbb{X}_2, d_2, \mu_2^*, \{\lambda_{2,k}\}, \{\phi_{2,k}\})$. We assume that a subset of “landmarks” $\mathbb{Y}_1 \subset \mathbb{X}_1$ is in a one-to-one correspondence with the corresponding subset of landmarks $\mathbb{Y}_2 \subset \mathbb{X}_2$, with some smoothness conditions on this correspondence as appropriate. To simplify notation, we will assume simply that $\mathbb{Y}_1 = \mathbb{Y}_2 = \mathbb{Y} \subseteq \mathbb{X}_1 \cap \mathbb{X}_2$, and that there is a positive measure ν^* supported on \mathbb{Y} such that $\nu^*(\mathbb{Y}) = 1$. The measure ν^* might well be discrete.

Our first objective in this section is to examine which functions on \mathbb{X}_2 can be lifted to functions on \mathbb{X}_1 via the landmarks \mathbb{Y} , and how the smoothness of such a lifted function changes from that of the original function. One example is the theory in Section 2.2, where the “lifted” function is Uf , and the smoothness does not change. A different model is the trace theorems for partial differential equations, where the smoothness of the trace/extension is typically different from the function one starts out with.

As a simple starting point, we replace $\langle \phi_{1,j}, \phi_{2,k} \rangle$ in the interaction term above by

$$\Gamma_{j,k} = \int_{\mathbb{Y}} \overline{\phi_{1,j}(y)} \phi_{2,k}(y) d\nu^*(y), \quad j, k = 0, 1, \dots \quad (3.1)$$

Let h be a smooth, low pass filter, and Γ be the matrix defined by (3.1). We define

$$\Phi_{n,\otimes}(\Xi_1, \Xi_2; h, x_1, x_2) = \sum_{j,k=0}^{\infty} h\left(\frac{\lambda_{1,j}}{n}\right) h\left(\frac{\lambda_{2,k}}{n}\right) \Gamma_{j,k} \phi_{1,j}(x_1) \overline{\phi_{2,k}(x_2)}, \quad x_1 \in \mathbb{X}_1, x_2 \in \mathbb{X}_2. \quad (3.2)$$

If $f \in L^1(\mathbb{X}_2, \mu_2^*) + L^\infty(\mathbb{X}_2, \mu_2^*)$, we define

$$\sigma_{n,\otimes}(\Xi_1, \Xi_2; h, f, x_1) = \int_{\mathbb{X}_2} f(x_2) \Phi_{n,\otimes}(\Xi_1, \Xi_2; h, x_1, x_2) d\mu_2^*(x_2), \quad x_1 \in \mathbb{X}_1, n > 0. \quad (3.3)$$

In the case when $\Xi_1 = (\mathbb{X}, d_1, \mu^*, \{\lambda_{1,j}\}, \{\phi_{1,j}\})$ and $\Xi_2 = (\mathbb{X}, d_2, \mu^*, \{\lambda_{2,k}\}, \{\phi_{2,k}\})$, $\mathbb{Y} = \mathbb{X}$, $\nu^* = \mu^*$, we have for each k ,

$$\sigma_{m,\otimes}(\Xi_1, \Xi_2; h, \phi_{2,k}, x_1) = h\left(\frac{\lambda_{2,k}}{m}\right) \sum_{j=0}^{\infty} h\left(\frac{\lambda_{1,j}}{m}\right) \Gamma_{j,k} \phi_{1,j}(x_1) = h\left(\frac{\lambda_{2,k}}{m}\right) \sigma_m(\Xi_1, \Xi_1; h, \phi_{2,k}, x_1).$$

It is not difficult to deduce from here, under suitable conditions as in [48], that $\lim_{m \rightarrow \infty} \sigma_{m,\otimes}(\Xi_1, \Xi_2; h, P) = P$ for every $P \in \Pi_\infty(\Xi_2)$. Moreover, the results in [48] show that if P is sufficiently smooth in Ξ_1 , then

$$\|\sigma_{2^{m+1},\otimes}(\Xi_1, \Xi_2; h, P) - \sigma_{2^m,\otimes}(\Xi_1, \Xi_2; h, P)\|_{\mathbb{X}_1, \mu_1^*, p} = \mathcal{O}(2^{-m\beta})$$

for some $\beta > 0$. In our more general case, this may or may not be true, depending upon the choice of \mathbb{Y} and ν^* . Nevertheless, in order to obtain a theorem about the change in smoothness of a function when “lifted” from Ξ_2 to Ξ_1 , we need to make a similar assumption ((3.5) below).

Theorem 3.1 *Let $\Xi_1 = (\mathbb{X}_1, d_1, \mu_1^*, \{\lambda_{1,k}\}, \{\phi_{1,k}\})$ and $\Xi_2 = (\mathbb{X}_2, d_2, \mu_2^*, \{\lambda_{2,k}\}, \{\phi_{2,k}\})$ be two admissible systems as in Definiton 2.1, each of which satisfies the Gaussian upper bound condition with exponents q_1, q_2 respectively, where $\{\phi_{1,j}\}$ and $\{\phi_{2,k}\}$ are orthonormalized with respect to μ_1^* , respectively, μ_2^* . Let \mathbb{Y} be a measurable (with respect to both μ_1^* and μ_2^*) subset of $\mathbb{X}_1 \cap \mathbb{X}_2$, ν^* be a probability measure on \mathbb{Y} , $\Gamma_{j,k}$ be defined as in (3.1). Let h be a sufficiently smooth low pass filter, and μ_1^*, μ_2^* satisfy the following regularity condition (cf. (2.13)):*

$$\mu_j^*(\mathbb{B}_j(x, r)) \leq cr^{q_j}, \quad x \in \mathbb{X}_j, \quad r > 0, \quad j = 1, 2. \quad (3.4)$$

Let $1 \leq p \leq \infty$. We assume further that there exists a $\beta > 0$ such that for any $m, n \geq 1$ and $P \in \Pi_n(\Xi_2)$,

$$\|\sigma_{2^{m+1},\otimes}(\Xi_1, \Xi_2; h, P) - \sigma_{2^m,\otimes}(\Xi_1, \Xi_2; h, P)\|_{\mathbb{X}_1, \mu_1^*, p} \leq c2^{-m\beta} n^{q_1+(q_2-q_1)/p} \|P\|_{\mathbb{X}_2, \mu_2^*, p}. \quad (3.5)$$

If $f \in L^p(\mathbb{X}_2, \mu_2^*)$, and

$$\sum_{m=0}^{\infty} 2^{m(q_1+(q_2-q_1)/p)} E_{2^m, p}(\Xi_2; f) < \infty, \quad (3.6)$$

then $\mathcal{E}_\otimes(f) = \mathcal{E}_\otimes(\Xi_1, \Xi_2; f) = \lim_{n \rightarrow \infty} \sigma_{2^n, \otimes}(\Xi_1, \Xi_2; h, f)$ exists in the sense of $L^p(\Xi_1, \mu_1^*)$, and

$$E_{2^n, p}(\Xi_1; \mathcal{E}_\otimes(f)) \leq c \left\{ \sum_{m=n}^{\infty} 2^{m(q_1+(q_2-q_1)/p)} E_{2^m, p}(\Xi_2; f) + 2^{n(q_1+(q_2-q_1)/p-\beta)} \|f\|_{\Xi_2, p} \right\}. \quad (3.7)$$

In particular, if $\beta > \gamma > q_1 + (q_2 - q_1)/p$, $0 < \rho \leq \infty$, and $f \in B_{p, \rho, \gamma}(\Xi_2)$, then $\mathcal{E}_\otimes(f) \in B_{p, \rho, \gamma - q_1 - (q_2 - q_1)/p}$.

We note that even if $\Xi_1 = \Xi_2$ in the above theorem, a straightforward application of the theorem leads to a loss of smoothness in f . This is a technicality due in part to the very general nature of \mathbb{Y} and ν^* , but also due to the definition of $\sigma_{n,\otimes}$. It is possible to eliminate this anomaly by defining the analogue of the various quantities in one system theory in our context in an abstract manner rather than assuming the identification of some elements of \mathbb{X}_1 with those in \mathbb{X}_2 via \mathbb{Y} .

In particular, we want to define a version of the Gaussian upper bound condition appropriate to two systems considered jointly. To motivate this discussion, we first make an observation, continuing the same scenario as in Theorem 3.1.

Example 3.1 Let $\Xi_1 = (\mathbb{X}_1, d_1, \mu_1^*, \{\lambda_{1,j}\}, \{\phi_{1,j}\})$ and $\Xi_2 = (\mathbb{X}_2, d_2, \mu_2^*, \{\lambda_{2,k}\}, \{\phi_{2,k}\})$ be two admissible systems as in Definiton 2.1, each of which satisfies the Gaussian upper bound condition with exponents q_1, q_2 respectively, where $\{\phi_{1,j}\}$ and $\{\phi_{2,k}\}$ are orthonormalized with respect to μ_1^* , respectively, μ_2^* . Let \mathbb{Y} be a measurable (with respect to both μ_1^* and μ_2^*) subset of $\mathbb{X}_1 \cap \mathbb{X}_2$, ν^* be a probability measure on \mathbb{Y} . We define, in this example only,

$$d_{1,2}(x_1, x_2) = \inf_{y \in \mathbb{Y}} (d_1(x_1, y) + d_2(y, x_2)). \quad (3.8)$$

Since $d_{1,2}$ is defined on $\mathbb{X}_1 \times \mathbb{X}_2$, it cannot be called a distance in the usual sense of the term, but it is clearly non-negative, satisfies the symmetry condition $d_{1,2}(x_1, x_2) = d_{2,1}(x_2, x_1)$, and the analogues of the triangle inequality:

$$\begin{aligned} d_{1,2}(x_1, x_2) &\leq c(d_1(x_1, x'_1) + d_{1,2}(x'_1, x_2)), & x_1, x'_1 \in \mathbb{X}_1, \quad x_2 \in \mathbb{X}_2, \\ d_{1,2}(x_1, x_2) &\leq c(d_{1,2}(x_1, x'_2) + d_2(x'_2, x_2)), & x_1 \in \mathbb{X}_1, \quad x_2, x'_2 \in \mathbb{X}_2. \end{aligned} \quad (3.9)$$

Next, we let $\ell_{j,k} = (\lambda_{1,j}^2 + \lambda_{2,k}^2)^{1/2}$, and

$$K_t(\Xi_1, \Xi_2; x_1, x_2) = \sum_{k,j=0}^{\infty} \exp(-\ell_{j,k}^2 t) \Gamma_{j,k} \phi_{1,j}(x_1) \overline{\phi_{2,k}(x_2)}. \quad (3.10)$$

An application of (2.12) shows that

$$\begin{aligned}
|K_t(\Xi_1, \Xi_2; x_1, x_2)| &= \left| \int_{\mathbb{Y}} K_t(\Xi_1; x_1, y) \overline{K_t(\Xi_2; x_2, y)} d\nu^*(y) \right| \\
&\leq \int_{\mathbb{Y}} |K_t(\Xi_1; x_1, y)| |K_t(\Xi_2; x_2, y)| d\nu^*(y) \\
&\leq ct^{-(q_1+q_2)/2} \int_{\mathbb{Y}} \exp\left(-c \frac{d_1(x_1, y)^2 + d_2(y, x_2)^2}{t}\right) d\nu^*(y) \\
&\leq ct^{-(q_1+q_2)/2} \exp\left(-c \frac{d_{1,2}(x_1, x_2)^2}{t}\right). \tag{3.11}
\end{aligned}$$

□

The next example is meant to illustrate the need for a more general set up which arises in a very routine extension of functions on the closed unit disc of \mathbb{R}^2 to the Euclidean sphere \mathbb{S}^2 in \mathbb{R}^3 .

Example 3.2 Let \mathbb{X}_1 be the Euclidean hemisphere

$$\mathbb{X}_1 = \{\mathbf{p}(\theta, \phi) = (\sin \theta \cos \phi, \sin \theta \sin \phi, \cos \theta) : \theta \in [0, \pi/2], \phi \in (-\pi, \pi]\},$$

d_1 be the geodesic distance on \mathbb{X}_1 , μ_1^* be the area measure, normalized to be a probability measure, $\lambda_{1,(j,\ell)} = \sqrt{(j+\ell)(j+\ell+1)}$, $j = 0, 1, \dots, \ell \in \mathbb{Z}$, and

$$\phi_{1,(j,\ell)}(\mathbf{p}(\theta, \phi)) = \sqrt{2} \sin^{|\ell|} \theta p_{2j+1}^{(|\ell|, |\ell|)}(\cos \theta) \exp(i\ell\phi),$$

where, in this example, $p_k^{(\alpha, \beta)}$ denotes the orthonormalized Jacobi polynomial of degree k ; i.e., a polynomial of degree k with positive leading coefficients, such that for integer $k, j \geq 0$,

$$\int_{-1}^1 p_k^{(\alpha, \beta)}(x) p_m^{(\alpha, \beta)}(x) (1-x)^\alpha (1+x)^\beta dx = \begin{cases} 1, & \text{if } k = m, \\ 0, & \text{otherwise.} \end{cases}$$

For \mathbb{X}_2 , we take the unit disc $\{(x, y) \in \mathbb{R}^2 : x^2 + y^2 \leq 1\}$, d_2 to be the Euclidean distance, μ_2^* to be the area measure, let $\mathbf{q}(\theta, \phi) = (\sin \theta \cos \phi, \sin \theta \sin \phi)$, $\theta \in [0, \pi/2]$, $\phi \in (-\pi, \pi]$, $\lambda_{2,(k,m)} = k|m| + k + |m|$, and define $\phi_{2,(k,m)}$ by

$$\phi_{2,(k,m)}(\mathbf{q}(\theta, \phi)) = \sqrt{\frac{2^{|m|+2}}{\pi}} \sin^{|m|} \theta \cos \theta p_k^{(|m|, 1)}(\cos(2\theta)) \exp(im\phi), \quad m \in \mathbb{Z}, k = 0, 1, \dots.$$

In this example, each member of the systems $\{\phi_{1,(j,\ell)}\}$, $\{\phi_{2,(k,m)}\}$ is equal to 0 on $\mathbb{X}_1 \cap \mathbb{X}_2$. So, the construction as in Example 3.1 does not work. There is, of course, an obvious one-to-one correspondence between each point $\mathbf{p}(\theta, \phi) \in \mathbb{X}_1$ and $\mathbf{q}(\theta, \phi) \in \mathbb{X}_2$, but since the measures are different, we are in a different setting from the one described in [15] as well. As an interaction term, we find it convenient to define for $j, k = 0, 1, \dots, \ell, m \in \mathbb{Z}$,

$$A_{(j,\ell),(k,m)} = \begin{cases} \sqrt{\pi} 2^{-|m|-5/4} a_{j,k}^{(m)}, & \text{if } \ell = m, 0 \leq k \leq j, \\ 0, & \text{otherwise,} \end{cases}$$

where

$$a_{j,k}^{(m)} = \int_{-1}^1 p_j^{(|m|, 1/2)}(x) p_k^{(|m|, 1)}(x) (1-x)^{|m|} (1+x) dx, \quad 0 \leq k \leq j, j = 0, 1, \dots, m \in \mathbb{Z}.$$

Explicit expressions for $a_{j,k}^{(m)}$ are known, cf. [1, Lecture 7]. Using the fact ([62, Theorem 4.1, Formula (4.3.4)]) that for $\alpha > -1$, $\theta \in [0, \pi/2]$,

$$p_{2j+1}^{(\alpha, \alpha)}(\cos \theta) = 2^{\alpha/2+3/4} \cos \theta p_j^{(\alpha, 1/2)}(\cos(2\theta)), \quad j = 0, 1, \dots, \tag{3.12}$$

we now deduce that

$$\sum_{k=0}^{\infty} \sum_{m \in \mathbb{Z}} A_{(j,\ell),(k,m)} \phi_{2,(k,m)}(\mathbf{q}(\theta_2, \phi_2)) = \phi_{1,(j,\ell)}(\mathbf{p}(\theta_2, \phi_2)).$$

Since $\phi_{1,(j,\ell)}$ are Dirichlet eigenfunctions of the Laplace–Beltrami operator on \mathbb{X}_1 corresponding to the eigenvalue $-\lambda_{1,(j,\ell)}^2$, it follows from the results in [43, 20] that

$$\begin{aligned} & \left| \sum_{j,k=0}^{\infty} \sum_{m,\ell \in \mathbb{Z}} \exp(-(j+\ell)(j+\ell+1)t) A_{(j,\ell),(k,m)} \phi_{1,(j,\ell)}(\mathbf{p}(\theta, \phi)) \overline{\phi_{2,(k,m)}(\mathbf{q}(\theta_2, \phi_2))} \right| \\ &= \left| \sum_{j=0}^{\infty} \sum_{\ell \in \mathbb{Z}} \exp(-(j+\ell)(j+\ell+1)t) \phi_{1,(j,\ell)}(\mathbf{p}(\theta_1, \phi_1)) \overline{\phi_{1,(j,\ell)}(\mathbf{p}(\theta_2, \phi_2))} \right| \\ &\leq c_1 t^{-1} \exp\left(-c_2 \frac{d_1((\mathbf{p}(\theta_1, \phi_1), (\mathbf{p}(\theta_2, \phi_2))^2)}{t}\right). \end{aligned}$$

We note finally that setting

$$d_{1,2}(\mathbf{p}(\theta_1, \phi_1), \mathbf{q}(\theta_2, \phi_2)) = d_1((\mathbf{p}(\theta_1, \phi_1), (\mathbf{p}(\theta_2, \phi_2))),$$

the function $d_{1,2} : \mathbb{X}_1 \times \mathbb{X}_2$ satisfies a triangle inequality analogous to (3.9). \square

Motivated by these examples, we make first the next definition of a joint distance.

Definition 3.1 *Let (\mathbb{X}_1, d_1) , (\mathbb{X}_2, d_2) be quasi-metric spaces. A function $d_{1,2} : \mathbb{X}_1 \times \mathbb{X}_2 \rightarrow [0, \infty)$, will be called a **joint distance** for the pair of quasi-metric spaces if the following triangle inequalities are satisfied.*

$$\begin{aligned} d_{1,2}(x_1, x_2) &\leq c(d_1(x_1, x'_1) + d_{1,2}(x'_1, x_2)), & x_1, x'_1 \in \mathbb{X}_1, \ x_2 \in \mathbb{X}_2, \\ d_{1,2}(x_1, x_2) &\leq c(d_{1,2}(x_1, x'_2) + d_2(x'_2, x_2)), & x_1 \in \mathbb{X}_1, \ x_2, x'_2 \in \mathbb{X}_2. \end{aligned} \quad (3.13)$$

We define

$$d_{2,1}(x_2, x_1) = d_{1,2}(x_1, x_2), \quad x_1 \in \mathbb{X}_1, \ x_2 \in \mathbb{X}_2, \quad (3.14)$$

For $r > 0$, $x_1 \in \mathbb{X}_1$, $x_2 \in \mathbb{X}_2$, we define the balls

$$\begin{aligned} \mathbb{B}_1(x_1, r) &= \{z \in \mathbb{X}_1 : d_1(x_1, z) \leq r\}, & \mathbb{B}_2(x_2, r) &= \{z \in \mathbb{X}_2 : d_2(x_2, z) \leq r\}. \\ \mathbb{B}_{1,2}(x_1, r) &= \{z \in \mathbb{X}_2 : d_{1,2}(x_1, z) \leq r\}, & \mathbb{B}_{2,1}(x_2, r) &= \{z \in \mathbb{X}_1 : d_{2,1}(x_2, z) \leq r\}. \end{aligned} \quad (3.15)$$

Next, we would like to allow more flexible connections than that provided by the matrix Γ as in (3.1). Example 3.2 demonstrates the need to do so. Another motivation is to incorporate some probabilistic transition from a point $x_1 \in \mathbb{X}_1$ to $x_2 \in \mathbb{X}_2$. These considerations prompt the following very general definition analogous to Definition 2.3.

Definition 3.2 *Let $\Xi_1 = (\mathbb{X}_1, d_1, \mu_1^*, \{\lambda_{1,k}\}, \{\phi_{1,k}\})$ and $\Xi_2 = (\mathbb{X}_2, d_2, \mu_2^*, \{\lambda_{2,k}\}, \{\phi_{2,k}\})$ be two admissible systems as in Definition 2.1, $d_{1,2}$ be the joint distance between the two systems. Let $\mathbf{A} = (A_{j,k})_{j,k=0}^{\infty}$, $\boldsymbol{\lambda} = (\ell_{j,k})_{j,k=0}^{\infty}$, where each $\ell_{j,k} \geq 0$, and for each $u > 0$, the set $\{(j,k) : \ell_{j,k} < u\}$ is finite. We define formally for $x_1 \in \mathbb{X}_1$, $x_2 \in \mathbb{X}_2$, $t > 0$, the **joint heat kernel** by*

$$\begin{aligned} K_t(\Xi_1, \Xi_2; x_1, x_2) &= K_t(\Xi_1, \Xi_2; \mathbf{A}, \boldsymbol{\lambda}; x_1, x_2) = \sum_{j,k=0}^{\infty} \exp(-\ell_{j,k}^2 t) A_{j,k} \phi_{1,j}(x_1) \overline{\phi_{2,k}(x_2)} \\ &= \lim_{n \rightarrow \infty} \sum_{j,k: \ell_{j,k} < n} \exp(-\ell_{j,k}^2 t) A_{j,k} \phi_{1,j}(x_1) \overline{\phi_{2,k}(x_2)}. \end{aligned} \quad (3.16)$$

We say that Ξ_1, Ξ_2 satisfy *The joint Gaussian upper bound condition* (with exponent $Q > 0$) if

$$\sum_{j,k: \ell_{j,k} < n} |A_{j,k} \phi_{1,j}(x_1) \overline{\phi_{2,k}(x_2)}| \leq cn^Q, \quad n \geq 2, \quad (3.17)$$

the limit in (3.16) exists for all $x_1 \in \mathbb{X}_1$, $x_2 \in \mathbb{X}_2$, and

$$|K_t(\Xi_1, \Xi_2; x_1, x_2)| \leq c_1 t^{-c} \exp\left(-c_2 \frac{d_{1,2}(x_1, x_2)^2}{t}\right), \quad x_1 \in \mathbb{X}_1, \ x_2 \in \mathbb{X}_2. \quad (3.18)$$

We will refer to \mathbf{A} as the **connection matrix**, its elements as **connection coefficients**, and the numbers $\ell_{j,k}$ as **joint eigenvalues**.

Remark 3.1 When $\Xi_1 = \Xi_2$, \mathbf{A} is the identity matrix, $\ell_{j,k} = \lambda_k \delta_{j,k}$, then (3.17) takes the form

$$\sum_{k: \lambda_k < n} \exp(-\lambda_k^2 t) |\phi_{1,k}(x_1)|^2 \leq cn^Q, \quad x_1 \in \mathbb{X}_1, \quad n \geq 2.$$

We have proved in [25, Proposition 4.1] that this is equivalent to the statement that

$$K_t(\Xi_1, \Xi_1; x_1, x_1) \leq ct^{-Q/2}, \quad 0 < t < 1.$$

Thus, the joint Gaussian upper bound condition in Definition 3.2 reduces to the Gaussian upper bound condition for a single system defined in Definition 2.3. \square

Example 3.3 In the manifold case as in Example 2.1, if ψ_k is a derivative of ϕ_k , then an integration by parts argument shows that $\{\psi_k\}$ is a Bessel system. It is known that under certain conditions, the systems $\Xi_1 = (\mathbb{X}, d, \mu^*, \{\lambda_k\}, \{\phi_k\})$, $\Xi_2 = (\mathbb{X}, d, \mu^*, \{\lambda_k\}, \{\psi_k\})$ satisfy the joint Gaussian upper bound condition ([32] and references therein). \square

In the remainder of this section, let $\Xi_1 = (\mathbb{X}_1, d_1, \mu_1^*, \{\lambda_{1,k}\}, \{\phi_{1,k}\})$ and $\Xi_2 = (\mathbb{X}_2, d_2, \mu_2^*, \{\lambda_{2,k}\}, \{\phi_{2,k}\})$ be two admissible systems satisfying the joint Gaussian upper bound condition with exponent $Q > 0$, and $d_{1,2}$, \mathbf{A} , $\boldsymbol{\lambda}$ be as in Definition 3.2.

Next, we define the analogues of the operators σ_n .

If h is a low pass filter, we define the kernel Φ_n for $x_1 \in \mathbb{X}_1$, $x_2 \in \mathbb{X}_2$ by

$$\Phi_n(\Xi_1, \Xi_2; h, x_1, x_2) = \Phi_n(\Xi_1, \Xi_2; \mathbf{A}, \boldsymbol{\lambda}; h, x_1, x_2) = \sum_{j,k=0}^{\infty} h\left(\frac{\ell_{j,k}}{n}\right) A_{j,k} \phi_{1,j}(x_1) \overline{\phi_{2,k}(x_2)}. \quad (3.19)$$

If $f \in L^1(\mathbb{X}_2, \mu_2^*) + L^\infty(\mathbb{X}_2, \mu_2^*)$, $x_1 \in \mathbb{X}_1$, $n > 0$, we define

$$\begin{aligned} \sigma_n(\Xi_1, \Xi_2; h, f, x_1) &= \sigma_n(\Xi_1, \Xi_2; \mathbf{A}, \boldsymbol{\lambda}; h, f, x_1) = \int_{\mathbb{X}_2} f(x_2) \Phi_n(\Xi_1, \Xi_2; h, x_1, x_2) d\mu_2^*(x_2) \\ &= \sum_{j,k=0}^{\infty} h\left(\frac{\ell_{j,k}}{n}\right) A_{j,k} \hat{f}(\Xi_2; k) \phi_{1,j}(x_1). \end{aligned} \quad (3.20)$$

Remark 3.2 In the case when the matrix \mathbf{A} is the identity matrix, $\{\phi_{1,j}\}$, $\{\phi_{2,k}\}$ are orthonormal systems, U is a unitary operator from $L^2(\mathbb{X}_2, \mu_2^*)$ into $L^2(\mathbb{X}_1, \mu_1^*)$ such that $\phi_{2,k} = U^* \phi_{1,k}$ for each k , then the kernels Φ_n and the corresponding operators σ_n correspond to these quantities defined in Section 2.2 in the case when $\mathbb{X}_1 = \mathbb{X}_2$, $\mu_1^* = \mu_2^*$. Because of the special relationship between the two systems, the results in Section 2.2 are valid in this more general context, assuming the Gaussian upper bound condition only for the system $\{\phi_{1,j}\}$, and not the system $\{\phi_{2,k}\}$. In the greater generality of this section, where there is no such natural connection between the systems, we need to assume a joint Gaussian upper bound condition for the following theorems to hold. \square

Our generalization of Theorem 3.1 is the following.

Theorem 3.2 Let $\Xi_1 = (\mathbb{X}_1, d_1, \mu_1^*, \{\lambda_{1,k}\}, \{\phi_{1,k}\})$ and $\Xi_2 = (\mathbb{X}_2, d_2, \mu_2^*, \{\lambda_{2,k}\}, \{\phi_{2,k}\})$ be two admissible systems satisfying the joint Gaussian upper bound condition with exponent $Q > 0$, and \mathbf{A} , $\boldsymbol{\lambda}$ as in Definition 3.2. Let h be a sufficiently smooth low pass filter, and μ_1^*, μ_2^* satisfy the **joint regularity condition**

$$\mu_1^*(\mathbb{B}_{2,1}(x_2, r)) \leq cr^{q_1}, \quad \mu_2^*(\mathbb{B}_{1,2}(x_1, r)) \leq cr^{q_2}, \quad x_1 \in \mathbb{X}_1, \quad x_2 \in \mathbb{X}_2, \quad r > 0, \quad (3.21)$$

for some constants $c, q_1, q_2 > 0$ independent of x_1 and r . Let $1 \leq p \leq \infty$, and $f \in L^p(\mathbb{X}_2, \mu_2^*)$. We assume further that there exists $\beta > Q - q_2 - (q_1 - q_2)/p$ such that for any $n \geq 1$ and $P \in \Pi_n(\Xi_2)$,

$$\|\sigma_{2^{m+1}}(\Xi_1, \Xi_2; \mathbf{A}, \boldsymbol{\lambda}; h, P) - \sigma_{2^m}(\Xi_1, \Xi_2; \mathbf{A}, \boldsymbol{\lambda}; h, P)\|_{\mathbb{X}_1, \mu_1^*, p} \leq c2^{-m\beta} n^{Q-q_2-(q_1-q_2)/p} \|P\|_{\mathbb{X}_2, \mu_2^*, p}, \quad m = 1, 2, \dots \quad (3.22)$$

If

$$\sum_{m=0}^{\infty} 2^{m(Q-q_2-(q_1-q_2)/p)} E_{2^m, p}(\Xi_2; f) < \infty, \quad (3.23)$$

then $\lim_{n \rightarrow \infty} \sigma_{2^n}(\Xi_1, \Xi_2; \mathbf{A}, \boldsymbol{\lambda}; h, f) = \mathcal{E}(f) = \mathcal{E}(\Xi_1, \Xi_2; \mathbf{A}, \boldsymbol{\lambda}; f)$ exists in the sense of $L^p(\mathbb{X}_1, \mu_1^*)$ and

$$\begin{aligned} & \|\mathcal{E}(f) - \sigma_{2^n}(\Xi_1, \Xi_2; \mathbf{A}, \boldsymbol{\lambda}; h, f)\|_{\mathbb{X}_1, \mu_1^*, p} \\ & \leq c \left\{ \sum_{m=n}^{\infty} 2^{m(Q-q_2-(q_1-q_2)/p)} E_{2^m, p}(\Xi_2; f) + 2^{n(Q-q_2-(q_1-q_2)/p-\beta)} \|f\|_{\mathbb{X}_2, \mu_2^*, p} \right\}. \end{aligned} \quad (3.24)$$

In particular, if $\alpha > 0$, and $\alpha \ell_{j,k} \geq \lambda_{1,j}$ for all $j, k = 0, 1, \dots$, then $\sigma_n(\Xi_1, \Xi_2; \mathbf{A}, \boldsymbol{\lambda}; h, f) \in \Pi_{\alpha n}(\Xi_1)$. Further, if $Q - q_2 - (q_1 - q_2)/p < \gamma < \beta$, and $f \in B_{p, \rho, \gamma}(\Xi_2)$ then $\mathcal{E}(f) \in B_{p, \rho, \gamma - Q + q_2 + (q_1 - q_2)/p}(\Xi_1)$.

Remark 3.3 In the case when $p = 2$, the condition (3.22) can be replaced by the condition that the spectral norms of dyadic sectors $(A_{j,k})_{2^{m-1} \leq \ell_{j,k} < 2^m, \lambda_{2,k} < n}$ are bounded from above by $c 2^{-m\beta} n^{Q-q_2-(q_1-q_2)/p}$. \square

Remark 3.4 In Theorem 3.2, we do not assume that $Q - q_2 - (q_1 - q_2)/p > 0$. Therefore, depending upon the sign of this expression, \mathcal{E} is a smoothing operator analogous to an integral operator, or smoothness preserving, or coarsening operation analogous to a differential operator. \square

Remark 3.5 The exponent in place of $Q - q_2 - (q_1 - q_2)/p$ in the “reverse” direction is $Q - q_1 - (q_2 - q_1)/p$. If $Q \neq (q_1 + q_2)/2$, this is not equal to the negative of the first exponent. This leads to the fact that it may not hold in general that under the conditions of Theorem 3.2, $f \in B_{p, \rho, \gamma}(\Xi_2)$ **if and only if** $\mathcal{E}(f) \in B_{p, \rho, \gamma - Q + q_2 + (q_2 - q_1)/p}(\Xi_1)$. Moreover, in order to obtain f back from $\mathcal{E}(f)$, one needs a construction with \mathbf{A} replaced by its “inverse” in some sense, and with some restrictions on $\boldsymbol{\lambda}$. We do not find it worthwhile to go into these details; this would lead only to a more complicated re-statement of Theorem 3.2 in two directions, without adding any new concepts. One sided analogues of Theorem 2.1 are quite easy to prove though, and we omit these details as well. \square

Remark 3.6 In the scenario considered in Example 3.2, $Q = q_1 = q_2 = 2$, and β can be chosen arbitrarily large. If $f : \mathbb{X}_2 \rightarrow \mathbb{C}$, The operator $\mathcal{E}(f)$ is the usual lifting to the sphere:

$$\mathcal{E}(f)(\mathbf{p}(\theta, \phi)) = f(\mathbf{q}(\theta, \phi)),$$

without requiring (3.23). Theorem 3.2 then confirms the well known fact that \mathcal{E} preserves the smoothness of the function f . \square

4 Proofs

4.1 Proofs of the theorems in Section 2.2 and Section 3.

PROOF OF THEOREM 2.1. As we mentioned earlier, this theorem is a direct consequence of the fact that

$$\hat{f}(\Xi'; k) = \int_{\mathbb{X}} f(y) \overline{\psi_k(y)} d\mu^*(y) = \int_{\mathbb{X}} f(y) \overline{(U^* \phi_k)(y)} d\mu^*(y) = \widehat{Uf}(\Xi; k),$$

so that

$$\sigma_n(\Xi, \Xi'; h, f) = \sigma_n(\Xi, \Xi; h, Uf).$$

The statements (a), (b), (c) of this theorem follow from [48, Theorem 2.1], and the statement (d) follows from [48, Theorem 3.1], applying these theorems with Uf in place of f . In [48], the conditions on the system Ξ were stated in terms of upper bounds on $\sum_{k: \lambda_k < n} |\phi_k(x)|^2$ and the finite speed of wave propagation. These were shown in [25] to be equivalent to the Gaussian upper bound condition. (See also Theorem 4.5 below.) \square

It is convenient to prove Theorem 3.2 first, before Theorem 3.1. An important ingredient of our proof is the following theorem.

Theorem 4.1 Let $\Xi_1 = (\mathbb{X}_1, d_1, \mu_1^*, \{\lambda_{1,k}\}, \{\phi_{1,k}\})$ and $\Xi_2 = (\mathbb{X}_2, d_2, \mu_2^*, \{\lambda_{2,k}\}, \{\phi_{2,k}\})$ be two admissible systems satisfying the joint Gaussian upper bound condition with exponent $Q > 0$, and $\mathbf{A}, \boldsymbol{\lambda}$ as in Definition 3.2. Let h be a sufficiently smooth low pass filter, and μ_1^*, μ_2^* satisfy the joint regularity condition (3.21). Let $1 \leq p \leq \infty$, and $f \in L^p(\mathbb{X}_2, \mu_2^*)$. Then for $n \geq 1$,

$$\|\sigma_n(\Xi_1, \Xi_2; h, f)\|_{\mathbb{X}_1, \mu_1^*, p} \leq cn^{Q-q_2-(q_1-q_2)/p} \|f\|_{\mathbb{X}_2, \mu_2^*, p}, \quad f \in L^p(\mathbb{X}_2, \mu_2^*). \quad (4.1)$$

Example 4.1 Taking $\Xi_1 = \Xi_2$, $Q = q_1 = q_2$, \mathbf{A} to be the identity matrix, $\lambda = \{\lambda_j \delta_{j,k}\}$, and $\{\phi_{1,j}\}$ to be an orthonormal system, we recover [48, Proposition 2.1]. This, in turn, leads to the other theorems in [48] used in the proof of Theorem 2.1. \square

Example 4.2 We consider an admissible system $\Xi_1 = (\mathbb{X}, d, \mu^*, \{\lambda_k\}, \{\phi_k\})$ as in Example 3.3 for \mathbb{X}_2 , define ψ_k as in that example, and let $\Xi_2 = (\mathbb{X}, d, \mu^*, \{\lambda_k\}, \{\psi_k\})$. Then Theorem 4.1 leads, in particular, to [25, Theorem 2.1]. \square

We will prove Theorem 4.1 by using the Tauberian theorem, Theorem 4.3, to be proved in Sub-section 4.2. For reasons of organizational convenience, we assume this theorem in this section, and note that if h is a sufficiently smooth low pass filter, then it can be extended to \mathbb{R} as an even function, that satisfies the conditions on H stipulated in Theorem 4.3.

In order to prove Theorem 4.1, we first prove a lemma.

Lemma 4.1 *If μ_2^* satisfies (3.21), then for $S > q_2$,*

$$\int_{\mathbb{X}_2} \frac{d\mu_2^*(x_2)}{\max(1, (nd_{1,2}(x_1, x_2)))^S} \leq cn^{-q_2}. \quad (4.2)$$

PROOF. In this proof only, let $\mathcal{A}_0 = \mathbb{B}_{1,2}(x_1, 1/n)$ and $\mathcal{A}_m = \mathbb{B}_{1,2}(x_1, 2^m/n) \setminus \mathbb{B}_{1,2}(x_1, 2^{m-1}/n)$, $m = 1, 2, \dots$. Then (3.21) shows that

$$\mu_2^*(\mathcal{A}_m) \leq c(2^m/n)^{q_2}, \quad m = 0, 1, \dots$$

Consequently,

$$\int_{\mathbb{X}_2} \frac{d\mu_2^*(x_2)}{\max(1, (nd_{1,2}(x_1, x_2)))^S} = \sum_{m=0}^{\infty} \int_{\mathcal{A}_m} \frac{d\mu_2^*(x_2)}{\max(1, (nd_{1,2}(x_1, x_2)))^S} \leq cn^{-q_2} \sum_{m=0}^{\infty} 2^{-m(S-q_2)} \leq cn^{-q_2}.$$

\square

In the remainder of this section, we assume that h is sufficiently smooth so as to satisfy the conditions (4.17)–(4.18) in Theorem 4.3 with $S > \max(q_1, q_2, Q + 1)$.

PROOF OF THEOREM 4.1. In this proof, we denote $\Phi_n(\Xi_1, \Xi_2; h, x_1, x_2)$ by $\Phi_n(x_1, x_2)$. For each $x_1 \in \mathbb{X}_1$ and $x_2 \in \mathbb{X}_2$, we define the measure μ on $[0, \infty)$ by

$$\mu([0, u)) = \mu(x_1, x_2; [0, u)) = \sum_{j,k: \ell_{j,k} < u} A_{j,k} \phi_{1,j}(x_1) \overline{\phi_{2,k}(x_2)}, \quad u \geq 0. \quad (4.3)$$

Using (3.17) and the fact that each of the functions $\phi_{1,j}$, $\phi_{2,k}$ are bounded on their domains, it follows that (cf. (4.15))

$$\|\mu\|_Q \leq c, \quad x_1 \in \mathbb{X}_1, \quad x_2 \in \mathbb{X}_2. \quad (4.4)$$

The estimate (3.18) is the same as (4.16) with $d_{1,2}(x_1, x_2)$ in place of r . Hence, the conditions of Theorem 4.3 are satisfied. Therefore, Theorem 4.3 shows that for each $x_1 \in \mathbb{X}_1$, $x_2 \in \mathbb{X}_2$,

$$\left| \int_0^\infty h(u/n) d\mu(x_1, x_2; \mu) \right| = \left| \sum_{j,k=0}^\infty h(\ell_{j,k}/n) A_{j,k} \phi_{1,j}(x_1) \overline{\phi_{2,k}(x_2)} \right| = |\Phi_n(x_1, x_2)| \leq c \frac{n^Q}{\max(1, (nd_{1,2}(x_1, x_2)))^S}. \quad (4.5)$$

Therefore, Lemma 4.1 (applied once with μ_1^* and once with μ_2^* as stated) leads to

$$\int_{\mathbb{X}_1} |\Phi_n(x_1, x_2)| d\mu_1^*(x_1) \leq cn^{Q-q_1}, \quad x_2 \in \mathbb{X}_2,$$

and

$$\int_{\mathbb{X}_2} |\Phi_n(x_1, x_2)| d\mu_2^*(x_2) \leq cn^{Q-q_2}, \quad x_1 \in \mathbb{X}_1.$$

These inequalities lead to (4.1) for $p = 1$ and $p = \infty$ respectively. The general case follows by an application of the Riesz–Thorin interpolation theorem. \square

PROOF OF THEOREM 3.2. In this proof, we write $\sigma_n(f)$ in place of $\sigma_n(\Xi_1, \Xi_2; h, f)$, and q in place of $Q - q_2 - (q_1 - q_2)/p$. Let $P \in \Pi_n(\Xi_2)$. Then (3.22) implies that

$$\sum_{m=0}^{\infty} \|\sigma_{2^{m+1}}(P) - \sigma_{2^m}(P)\|_{\Xi_1, \mu_1^*, p} \leq c \sum_{m=0}^{\infty} 2^{-m\beta} n^q \|P\|_{\Xi_2, \mu_2^*, p} \leq cn^q \|P\|_{\Xi_2, \mu_2^*, p}. \quad (4.6)$$

Therefore, there exists $\mathcal{E}(P) \in L^p(\mathbb{X}_1, \mu_1^*)$ such that

$$\mathcal{E}(P) = \sigma_1(P) + \sum_{m=0}^{\infty} (\sigma_{2^{m+1}}(P) - \sigma_{2^m}(P)), \quad (4.7)$$

where the series converges in the sense of $L^p(\mathbb{X}_1, \mu_1^*)$. Moreover, \mathcal{E} is a linear operation on $\Pi_{\infty}(\Xi_2)$. We conclude from (4.1) and (4.6) that for $m = 1, 2, \dots$,

$$\|\mathcal{E}(P)\|_{\mathbb{X}_1, \mu_1^*, p} \leq cn^q \|P\|_{\mathbb{X}_2, \mu_2^*, p}, \quad \|\mathcal{E}(P) - \sigma_{2^m}(P)\|_{\mathbb{X}_1, \mu_1^*, p} \leq cn^q 2^{-m\beta} \|P\|_{\mathbb{X}_2, \mu_2^*, p}, \quad P \in \Pi_n(\Xi_2). \quad (4.8)$$

Next, let $f \in L^p(\mathbb{X}_2, \mu_2^*)$. We find a sequence $\{P_m\}$ such that each $P_m \in \Pi_{2^m}(\Xi_2)$, and

$$\|f - P_m\|_{\mathbb{X}_2, \mu_2^*, p} \leq cE_{2^m, p}(\Xi_2; f), \quad m = 0, 1, \dots$$

Then $P_{m+1} - P_m \in \Pi_{2^{m+1}}(\Xi_2)$, and a simple application of triangle inequality shows that

$$\|P_{m+1} - P_m\|_{\mathbb{X}_2, \mu_2^*, p} \leq cE_{2^m, p}(\Xi_2; f).$$

Therefore, if (3.23) holds, then (4.8) implies that

$$\|\mathcal{E}(P_0)\|_{\mathbb{X}_1, \mu_1^*, p} + \sum_{m=0}^{\infty} \|\mathcal{E}(P_{m+1}) - \mathcal{E}(P_m)\|_{\mathbb{X}_1, \mu_1^*, p} \leq c \left\{ \|P_0\|_{\mathbb{X}_2, \mu_2^*, p} + \sum_{m=0}^{\infty} 2^{mq} E_{2^m, p}(\Xi_2; f) \right\} < \infty. \quad (4.9)$$

Consequently, there exists $\mathcal{E}(f) \in L^p(\mathbb{X}_1, \mu_1^*)$ such that

$$\mathcal{E}(f) = \mathcal{E}(P_0) + \sum_{m=0}^{\infty} (\mathcal{E}(P_{m+1}) - \mathcal{E}(P_m)),$$

with the series converging in the sense of $L^p(\mathbb{X}_1, \mu_1^*)$. Moreover,

$$\|\mathcal{E}(f) - \mathcal{E}(P_n)\|_{\mathbb{X}_1, \mu_1^*, p} \leq c \sum_{m=n}^{\infty} 2^{mq} E_{2^m, p}(\Xi_2; f). \quad (4.10)$$

Using this estimate, (4.1), and (4.8), we conclude that

$$\begin{aligned} \|\mathcal{E}(f) - \sigma_{2^n}(f)\|_{\mathbb{X}_1, \mu_1^*, p} &\leq \|\mathcal{E}(f) - \mathcal{E}(P_n)\|_{\mathbb{X}_1, \mu_1^*, p} + \|\mathcal{E}(P_n) - \sigma_{2^n}(P_n)\|_{\mathbb{X}_1, \mu_1^*, p} + \|\sigma_{2^n}(f - P_n)\|_{\mathbb{X}_1, \mu_1^*, p} \\ &\leq c \sum_{m=n}^{\infty} 2^{mq} E_{2^m, p}(\Xi_2; f) + c2^{n(q-\beta)} \|P_n\|_{\mathbb{X}_2, \mu_2^*, p} + c2^{nq} \|f - P_n\|_{\mathbb{X}_2, \mu_2^*, p} \\ &\leq c \sum_{m=n}^{\infty} 2^{mq} E_{2^m, p}(\Xi_2; f) + c2^{n(q-\beta)} \|f\|_{\mathbb{X}_2, \mu_2^*, p}. \end{aligned}$$

This proves (3.24), and in particular, the fact that $\mathcal{E}(f) = \lim_{n \rightarrow \infty} \sigma_{2^n}(f)$ in the sense of $L^p(\mathbb{X}_1, \mu_1^*)$.

We observe that the sum expression in (3.20) involves non-zero terms only for j, k for which $\ell_{j,k} < n$. If $\alpha\ell_{j,k} \geq \lambda_{1,j}$ for $j, k = 0, 1, \dots$, then the only values of j for which the summand in (3.20) is non-zero are those for which $\lambda_{1,j} < \alpha n$. Therefore, $\sigma_n(f) \in \Pi_{\alpha n}(\Xi_1)$. Therefore, (3.24) implies that

$$E_{\alpha 2^n, p}(\Xi_1; f) \leq c \left\{ \sum_{m=n}^{\infty} 2^{mq} E_{2^m, p}(\Xi_2; f) + 2^{n(q-\beta)} \|f\|_{\mathbb{X}_2, \mu_2^*, p} \right\}. \quad (4.11)$$

If $f \in B_{p, \rho, \gamma}(\Xi_2)$, then by definition the sequence $\{2^{mq} E_{2^m, p}(\Xi_2; f)\}$ is in $\mathbf{b}_{\rho, \gamma-q}$. The discrete Hardy inequality [21, Lemma 3.4, p. 27] implies that the sequence

$$\left\{ \sum_{m=n}^{\infty} 2^{mq} E_{2^m, p}(\Xi_2; f) \right\}_{n=0}^{\infty} \in \mathbf{b}_{\rho, \gamma-q}.$$

Since $\beta > \gamma > q$, then $\{2^{n(q-\beta)}\}_{n=0}^{\infty} \in \mathbf{b}_{\rho, \gamma-q}$ as well. Hence, (4.11) leads to the fact that $\mathcal{E}(f) \in B_{p, \rho, \gamma-q}$. \square

Next, we turn to the proof of Theorem 3.1. The analogue of Theorem 4.1 in this case is the following theorem.

Theorem 4.2 Let $\Xi_1 = (\mathbb{X}_1, d_1, \mu_1^*, \{\lambda_{1,k}\}, \{\phi_{1,k}\})$ and $\Xi_2 = (\mathbb{X}_2, d_2, \mu_2^*, \{\lambda_{2,k}\}, \{\phi_{2,k}\})$ be two admissible systems as in Definition 2.1, each of which satisfies the Gaussian upper bound condition with exponents q_1, q_2 respectively, where $\{\phi_{1,j}\}$ and $\{\phi_{2,k}\}$ are orthonormalized with respect to μ_1^* , respectively, μ_2^* . Let \mathbb{Y} be a measurable (with respect to both μ_1^* and μ_2^*) subset of $\mathbb{X}_1 \cap \mathbb{X}_2$, ν^* be a probability measure on \mathbb{Y} , $\Gamma_{j,k}$ be defined as in (3.1). We assume further that regularity condition (3.4) holds. Let $1 \leq p \leq \infty$, $f \in L^p(\mathbb{X}_2, \mu_2^*)$, and h be a sufficiently smooth low pass filter. Then

$$\|\sigma_{n,\otimes}(\Xi_1, \Xi_2; h, f)\|_{\mathbb{X}_1, \mu_1^*, p} \leq cn^{q_1+(q_2-q_1)/p} \|f\|_{\mathbb{X}_2, \mu_2^*, p}. \quad (4.12)$$

PROOF. In this proof, we write for $j = 1, 2$, $x, y \in \mathbb{X}_j$, $f \in L^1(\mathbb{X}_j, \mu_j^*)$,

$$\Phi_{n,j}(x, y) = \sum_{k=0}^{\infty} h\left(\frac{\lambda_{j,k}}{n}\right) \phi_{j,k}(x) \overline{\phi_{j,k}(y)}, \quad \sigma_{n,j}(f)(x) = \int_{\mathbb{X}} \Phi_{n,j}(x, y) f(y) d\mu_j^*(y). \quad (4.13)$$

We observe from the definitions that

$$\Phi_{n,\otimes}(\Xi_1, \Xi_2; h, x_1, x_2) = \int_{\mathbb{Y}} \Phi_{n,1}(x_1, z) \overline{\Phi_{n,2}(x_2, z)} d\nu^*(z). \quad (4.14)$$

As in the proof of Lemma 4.1, we see using (3.4) that

$$\int_{\mathbb{X}_2} |\Phi_{n,\otimes}(\Xi_1, \Xi_2; h, x_1, x_2)| d\mu_2^*(x_2) \leq \int_{\mathbb{Y}} |\Phi_{n,1}(x_1, z)| \int_{\mathbb{X}_2} |\Phi_{n,2}(x_2, z)| d\mu_2^*(x_2) d\nu^*(z) \leq cn^{q_1},$$

and an analogous inequality holds with integrals on \mathbb{X}_1 instead. Hence, using the Riesz–Thorin theorem as in the proof of Theorem 4.1, we arrive at (4.12). \square

PROOF OF THEOREM 3.1. This theorem proved exactly in the same way as in the proof of Theorem 4.1, using Theorem 4.2 in place of Theorem 4.1. The condition on the joint eigen-values in Theorem 3.2 is superfluous in this context; it is clear that if $P \in \Pi_n(\Xi_2)$, then $\sigma_{n,\otimes}(\Xi_1, \Xi_2; h, P) \in \Pi_n(\Xi_1)$. We omit the remaining details of this proof. \square

4.2 Tauberian theorem

We recall that if μ is an extended complex valued Borel measure on \mathbb{R} , then its total variation measure is defined for a Borel set B by

$$|\mu|(B) = \sup \sum |\mu(B_k)|,$$

where the sum is over a partition $\{B_k\}$ of B comprising Borel sets, and the supremum is over all such partitions.

A measure μ on \mathbb{R} is called an even measure if $\mu((-u, u)) = 2\mu([0, u))$ for all $u > 0$, and $\mu(\{0\}) = 0$. If μ is an extended complex valued measure on $[0, \infty)$, and $\mu(\{0\}) = 0$, we define a measure μ_e on \mathbb{R} by

$$\mu_e(B) = \mu(\{|x| : x \in B\}),$$

and observe that μ_e is an even measure such that $\mu_e(B) = \mu(B)$ for $B \subset [0, \infty)$. In the sequel, we will assume that all measures on $[0, \infty)$ which do not associate a nonzero mass with the point 0 are extended in this way, and will abuse the notation μ also to denote the measure μ_e . In the sequel, the phrase “measure on \mathbb{R} ” will refer to an extended complex valued Borel measure having bounded total variation on compact intervals in \mathbb{R} , and similarly for measures on $[0, \infty)$.

Our main Tauberian theorem is the following.

Theorem 4.3 Let μ be an extended complex valued measure on $[0, \infty)$, and $\mu(\{0\}) = 0$. We assume that there exist $Q, r > 0$, such that each of the following conditions are satisfied.

1.

$$\|\mu\|_Q := \sup_{u \in [0, \infty)} \frac{|\mu|([0, u))}{(u+2)^Q} < \infty, \quad (4.15)$$

2. There are constants $c, C > 0$, such that

$$\left| \int_{\mathbb{R}} \exp(-u^2 t) d\mu(u) \right| \leq c_1 t^{-C} \exp(-r^2/t) \|\mu\|_Q, \quad 0 < t \leq 1. \quad (4.16)$$

Let $H : [0, \infty) \rightarrow \mathbb{R}$, $S > Q + 1$ be an integer, and suppose that there exists a measure $H^{[S]}$ such that

$$H(u) = \int_0^\infty (v^2 - u^2)_+^S dH^{[S]}(v), \quad u \in \mathbb{R}, \quad (4.17)$$

and

$$V_{Q,S}(H) = \max \left(\int_0^\infty (v+2)^Q v^{2S} d|H^{[S]}|(v), \int_0^\infty (v+2)^Q v^S d|H^{[S]}|(v) \right) < \infty. \quad (4.18)$$

Then for $n \geq 1$,

$$\left| \int_0^\infty H(u/n) d\mu(u) \right| \leq c \frac{n^Q}{\max(1, (nr)^S)} V_{Q,S}(H) \|\mu\|_Q. \quad (4.19)$$

The proof of Theorem 4.3 will be given in a number of steps. First, we will show that the conclusions of the theorem are valid if they are valid when the function H is replaced by a special function. The next step is to show that the theorem is valid if the distributional Fourier transform of μ is supported on $(-\infty, -2r] \cup [2r, \infty)$. The final step is to prove that the assumptions in Theorem 4.3 imply this property for the distributional Fourier transform.

STEP 1.

Let

$$x_+^r = \begin{cases} x^r, & \text{if } x > 0, \\ 0, & \text{if } x \leq 0. \end{cases} \quad (4.20)$$

The Bochner–Riesz summability method applied to μ is defined by

$$\mathfrak{R}_{S;n}(\mu) := \int_{\mathbb{R}} \left(1 - \frac{u^2}{n^2}\right)_+^S d\mu(u). \quad (4.21)$$

Lemma 4.2 *Let $S > Q > 0$, and μ be an even measure satisfying (4.15) (and $\mu(\{0\}) = 0$). If*

$$\mathfrak{R}_{S;n}(\mu) \leq c \frac{(n+2)^Q}{\max(1, (nr)^S)} \|\mu\|_Q, \quad n > 0, \quad (4.22)$$

then (4.19) holds for all H satisfying (4.17), (4.18).

PROOF. Without loss of generality, we may assume that $\|\mu\|_Q = 1$. Let $n \geq 1$. Assuming that a change of the order of integration below is valid, we obtain from (4.17) that

$$\int_0^\infty H(u/n) d\mu(u) = \int_0^\infty \int_0^\infty (v^2 - (u/n)^2)_+^S dH^{[S]}(v) d\mu(u) = (1/2) \int_0^\infty \mathfrak{R}_{S;nv}(\mu) v^{2S} dH^{[S]}(v). \quad (4.23)$$

We observe that for $n \geq 1$, $v \geq 0$,

$$(nv + 2) \leq n(v + 2).$$

Therefore, the assumption (4.22) implies that if $nr > 1$ then

$$\begin{aligned} \left| \int_0^\infty H(u/n) d\mu(u) \right| &\leq cn^Q \int_0^\infty \frac{(v+2)^Q v^{2S} d|H^{[S]}|(v)}{\max(1, (nr)^S)} \\ &= cn^Q \left\{ \int_0^{1/(nr)} (v+2)^Q v^{2S} d|H^{[S]}|(v) + (nr)^{-S} \int_{1/(nr)}^\infty (v+2)^Q v^S d|H^{[S]}|(v) \right\} \\ &= \frac{cn^Q}{(nr)^S} \int_0^\infty (v+2)^Q v^S d|H^{[S]}|(v). \end{aligned} \quad (4.24)$$

If $nr \leq 1$, then (4.23) and (4.22) together imply that

$$\left| \int_0^\infty H(u/n) d\mu(u) \right| \leq cn^Q \int_0^\infty (v+2)^Q v^{2S} d|H^{[S]}|(v). \quad (4.25)$$

Using the condition (4.18), we deduce (4.19) immediately from (4.24), (4.25).

It remains to verify that the interchange of order of integration in (4.23) is justified. Using (4.15) and (4.18), we observe that

$$\begin{aligned} \int_0^\infty \int_0^\infty (v^2 - (u/n)^2)_+^S d|\mu|(u) d|H^{[S]}|(v) &= \int_0^\infty \int_0^{nv} (v^2 - (u/n)^2)_+^S d|\mu|(u) d|H^{[S]}|(v) \\ &\leq \int_0^\infty v^{2S} \int_0^{nv} d|\mu|(u) d|H^{[S]}|(v) \leq n^Q \int_0^\infty (v+2)^Q v^{2S} d|H^{[S]}|(v) < \infty. \end{aligned}$$

This justifies the change of the order of integration, and completes the proof. \square

Remark 4.1 The verification for the interchange of order of integration shows in particular that if $V_{Q,S}(H) < \infty$, then $\int_0^\infty |H(u/n)| d|\mu|(u) < \infty$ for all $n \geq 1$. \square

STEP 2.

Here we work with the support of the distributional Fourier transform of μ . We need first to recall a few facts about distributions and their Fourier transforms, as well as certain facts from measure theory, mainly to establish notation, and explain our abuse thereof.

A function $\phi : \mathbb{R} \rightarrow \mathbb{R}$ is called a rapidly decreasing function (or test function) if it is infinitely differentiable and $|x|^k \phi^{(j)}(x) \rightarrow 0$ as $|x| \rightarrow \infty$ for all non-negative integers k, j . The space of all test functions is denoted by \mathcal{S} , and is a locally convex linear space. Its (continuous) dual is denoted by \mathcal{S}' , and a member of this dual is called a (tempered) distribution. If $\phi \in \mathcal{S}$ its Fourier transform $\hat{\phi}$ and inverse Fourier transform $\tilde{\phi}$ are defined by

$$\hat{\phi}(x) = \frac{1}{2\pi} \int_{\mathbb{R}} \phi(y) \exp(-ixy) dy, \quad \tilde{\phi}(x) = \int_{\mathbb{R}} \phi(y) \exp(ixy) dy, \quad x \in \mathbb{R}. \quad (4.26)$$

Both $\hat{\phi}$ and $\tilde{\phi}$ are test functions. Likewise, if $\mu \in \mathcal{S}'$, its Fourier transform $\hat{\mu}$ and inverse Fourier transform $\tilde{\mu}$ are defined by

$$\hat{\mu}(\phi) = \mu(\hat{\phi}), \quad \tilde{\mu}(\phi) = \mu(\tilde{\phi}), \quad \phi \in \mathcal{S}. \quad (4.27)$$

Clearly, both $\hat{\mu}$ and $\tilde{\mu}$ are distributions. One has the inversion formulas

$$\hat{\hat{\phi}} = \tilde{\tilde{\phi}} = \phi, \quad \hat{\tilde{\mu}} = \tilde{\hat{\mu}} = \mu, \quad \phi \in \mathcal{S}, \quad \mu \in \mathcal{S}'.$$

We refer the reader to [58, 59] for further information regarding Fourier transforms, and their properties.

If μ is a signed, extended complex valued, even, Borel measure on \mathbb{R} , we will identify it with the corresponding function on \mathbb{R} , defined by $u \mapsto \mu([0, |u|])$, $u \in \mathbb{R}$, and abuse the notation to denote both the measure and the function by μ . The total variation measure $|\mu|$ of μ is also identified with the total variation function of the corresponding function on \mathbb{R} , and with the same abuse of notation. If there exists $Q \geq 0$ such that $|\mu|((-u, u)) \leq cu^Q$, $u > 0$, then μ also defines an element of \mathcal{S}' by the formula

$$\mu(\phi) = \int_{\mathbb{R}} \phi d\mu, \quad |\mu|(\phi) = \int_{\mathbb{R}} \phi d|\mu|, \quad \phi \in \mathcal{S}. \quad (4.28)$$

One advantage of the identification between the measure, the function, and the distribution is that one can use such theorems from measure theory as Fubini's theorem, dominated convergence theorem, etc. in many cases involving integration of test functions with respect to extended complex valued measures satisfying the condition just described. In particular, it is often useful to integrate by parts in (4.28) to observe that

$$\mu(\phi) = - \int_{\mathbb{R}} \phi'(u) \mu(u) du, \quad |\mu|(\phi) = - \int_{\mathbb{R}} \phi'(u) |\mu|(u) du, \quad \phi \in \mathcal{S}. \quad (4.29)$$

It is unfortunately customary to denote the measure, the function, and the distribution all by the same symbol, with the corresponding abuse of notations regarding the Fourier transforms, e.g.,

$$\hat{\mu}(\phi) = \frac{1}{2\pi} \int_{\mathbb{R}} \hat{\phi} d\mu, \quad \phi \in \mathcal{S},$$

and we will follow this convention, with clarifying remarks in case we feel a cause for confusion. The support of a function $f : \mathbb{R} \rightarrow \mathbb{C}$, denoted by $\text{supp}(f)$, is the closure of the set $\{x \in \mathbb{R} : f(x) \neq 0\}$, and f is said to be supported on any super-set of this set. The support of a measure μ on \mathbb{R} that can be viewed as a distribution is the closure of the set $A \subset \mathbb{R}$ such that $\int_{\mathbb{R}} \phi d\mu = 0$ for every $\phi \in \mathcal{S}$, supported on $\mathbb{R} \setminus A$.

We are now ready to state our auxiliary Tauberian theorem.

Theorem 4.4 *Let μ be an extended complex valued, even measure supported on \mathbb{R} (with $\mu(\{0\}) = 0$). We assume that there exist $Q, r > 0$, such that (4.15) holds, and the (distributional) Fourier transform $\hat{\mu}$ of μ is supported on $(-\infty, -2r] \cup [2r, \infty)$. Then the conclusion of Theorem 4.3 holds.*

In order to prove this theorem, we will show that the assumptions imply (4.22). First, we need a technical lemma.

Lemma 4.3 *Let $S \geq 2$ be an integer. For every $Y > 1$, there exists an even function $H_Y \in \mathcal{S}$ such that*

$$\widehat{H_Y}(x) = 0, \quad |x| \geq Y, \quad (4.30)$$

$$\left| \frac{d^k}{du^k} (1 - u^2)_+^S - H_Y^{(k)}(u) \right| \leq cY^{k-S}, \quad u \in \mathbb{R}, \quad k = 0, 1, \dots, S-1, \quad (4.31)$$

and for any integer $N \geq 1$,

$$|H_Y^{(k)}(u)| \leq c(N)(Y|u|)^{-N}, \quad |u| \geq 2, \quad k = 0, 1, \dots, S-1. \quad (4.32)$$

PROOF. The proof is by a direct construction. In this proof, we denote $F(u) = (1 - u^2)_+^S$. Since $S \geq 2$, F has a first derivative having a bounded total variation on \mathbb{R} . Therefore, the Fourier inversion formula holds for F . The Fourier transform of F is given in terms of the Bessel functions [65, p. 53] by

$$\begin{aligned} \hat{F}(x) &= \frac{S!2^{S-1/2}}{\sqrt{\pi}} \frac{J_{S+1/2}(x)}{x^{S+1/2}} \\ &= \frac{S!2^{S-1}}{\pi} \frac{1}{x^{S+1}} \left\{ e^{ix} \sum_{k=0}^S \frac{i^{k-S-1}(S+k)!}{k!(S-k)!(2x)^k} + e^{-ix} \sum_{k=0}^S \frac{(-i)^{k-S-1}(S+k)!}{k!(S-k)!(2x)^k} \right\}. \end{aligned} \quad (4.33)$$

Since F is supported on $[-1, 1]$, \hat{F} is an entire function of finite exponential type 1. In this proof, let h be an infinitely differentiable low pass filter. (Necessarily, h is a test function.) We define H_Y by

$$\widehat{H_Y}(x) = \hat{F}(x)h(x/Y), \quad x \in \mathbb{R}, \quad (4.34)$$

or equivalently,

$$H_Y(u) = \frac{Y}{2\pi} \int_{\mathbb{R}} F(v) \tilde{h}(Y(u-v)) dv = \frac{Y}{2\pi} \int_{-1}^1 F(v) \tilde{h}(Y(u-v)) dv, \quad u \in \mathbb{R}. \quad (4.35)$$

It is not difficult to verify from the definitions that $\widehat{H_Y} \in \mathcal{S}$, and hence, $H_Y \in \mathcal{S}$ also. The equation (4.30) is also clear from the definition of $\widehat{H_Y}$.

In this proof, let $g(u) = h(u) - h(2u)$, $u \in \mathbb{R}$, and for every integer $n \geq 1$, $G_{n,Y}$ be defined by

$$\widehat{G_{n,Y}}(x) = \hat{F}(x)g(x/(2^n Y)), \quad G_{n,Y}(u) = \frac{2^n Y}{2\pi} \int_{\mathbb{R}} F(v) \tilde{g}(2^n Y(u-v)) dv, \quad x, u \in \mathbb{R}. \quad (4.36)$$

For every integer $n \geq 1$, $g(x/(2^n Y)) = 0$ if $|x| \leq 2^{n-2}Y$ or $|x| \geq 2^n Y$. Therefore, the second equation in (4.33) shows that for $0 \leq k \leq S-1$,

$$\left| \widehat{G_{n,Y}^{(k)}}(x) \right| = |x|^k |\widehat{G_{n,Y}}(x)| \leq c(2^n Y)^{k-S-1}, \quad x \in \mathbb{R}. \quad (4.37)$$

Since $\widehat{G_{n,Y}^{(k)}}$ is supported on $[-2^n Y, -2^{n-2}Y] \cup [2^{n-2}Y, 2^n Y]$, the Fourier inversion formula implies that

$$\left| G_{n,Y}^{(k)}(u) \right| \leq \left| \int_{2^{n-2}Y \leq |x| \leq 2^n Y} e^{iux} \widehat{G_{n,Y}^{(k)}}(x) dx \right| \leq c(2^n Y)^{k-S}, \quad u \in \mathbb{R}, \quad k = 0, 1, \dots, S-1.$$

Consequently,

$$\sum_{n=1}^{\infty} \left| G_{n,Y}^{(k)}(u) \right| \leq cY^{k-S}, \quad u \in \mathbb{R}, \quad k = 0, 1, \dots, S-1, \quad (4.38)$$

and hence,

$$F^{(k)}(u) = H_Y^{(k)}(u) + \sum_{n=1}^{\infty} G_{n,Y}^{(k)}(u), \quad u \in \mathbb{R}, \quad k = 0, 1, \dots, S-1, \quad (4.39)$$

where the series converges uniformly and absolutely on \mathbb{R} . Moreover, (4.31) is clear from (4.39) and (4.38).

Next, we observe that for $|u| \geq 2$, $|v| \leq 1$, $|u-v| \geq (1/2)|u|$. Since \tilde{h} is a test function, (4.35) implies that for $|u| \geq 2$,

$$|H_Y^{(k)}(u)| \leq cY^{k+1} \int_{-1}^1 |F(v)| |\tilde{h}^{(k)}(Y(u-v))| dv \leq c(N)Y^{k+1}(Y|u|)^{-k-1-N} \int_{-1}^1 |F(v)| dv \leq c(N)(Y|u|)^{-N}.$$

□

We are now ready to prove Theorem 4.4.

PROOF OF THEOREM 4.4. In this proof, we may assume without loss of generality that $\|\mu\|_Q = 1$. We will prove (4.22). This estimate is clear immediately from (4.15) if $nr \leq 1$. In this proof, we assume $nr > 1$, let $Y = 2nr$, and H_Y denote the function as in Lemma 4.3. Then $\widehat{H_Y(\cdot/n)}$ is supported on $(-2r, 2r)$. Since $\hat{\mu}$ is supported on $(-\infty, -2r] \cup [2r, \infty)$,

$$\int_{\mathbb{R}} H_Y(u/n) d\mu(u) = 0. \quad (4.40)$$

Using (4.15), and (4.31) with $k = 0$, we deduce that

$$\left| \int_{-2n}^{2n} \left\{ \left(1 - \frac{u^2}{n^2}\right)_+^S - H_Y(u/n) \right\} d\mu(u) \right| \leq c \frac{(n+2)^Q}{(nr)^S}. \quad (4.41)$$

If $|u| \geq 2n$, then

$$\left(1 - \frac{u^2}{n^2}\right)_+^S - H_Y(u/n) = -H_Y(u/n).$$

Using integration by parts and (4.32) with $N = S > Q + 1$, we obtain

$$\begin{aligned} & \left| \int_{|u| \geq 2n} \left\{ \left(1 - \frac{u^2}{n^2}\right)_+^S - H_Y(u/n) \right\} d\mu(u) \right| \\ &= \left| \int_{|u| \geq 2n} H_Y(u/n) d\mu(u) \right| \leq |\mu|(2n)H_Y(2) + \frac{1}{n} \left| \int_{|u| \geq 2n} H_Y'(u/n) \mu(u) du \right| \\ &\leq c(n+2)^Q (nr)^{-S} + \frac{1}{nr^S} \int_{|u| \geq 2n} u^{Q-S} du \\ &\leq cn^Q (nr)^{-S}. \end{aligned} \quad (4.42)$$

The estimate (4.22) follows from (4.40), (4.41), (4.42). In turn, Lemma 4.2 now implies Theorem 4.4. □

STEP 3.

We will prove that the assumptions on the measure in Theorem 4.3 imply those in Theorem 4.4; i.e., we prove the following theorem using some of the ideas in [60, 25].

Theorem 4.5 *Let μ be an even measure satisfying (4.15). If there exist $r, C > 0$ such that (4.16) is satisfied. Then the (distributional) Fourier transform $\hat{\mu}$ of μ is supported on $(-\infty, -2r] \cup [2r, \infty)$.*

The proof of this theorem is fairly complicated, and is organized in the form of three preparatory lemmas.

Lemma 4.4 *Let μ be an even measure satisfying (4.15), and $R \geq 4$. If ϕ is a test function, and h is an infinitely differentiable low pass filter, then for every integer $N \geq 2$,*

$$\left| \int_{\mathbb{R}} \phi(u) h(u/R) d\mu(u) - \int_{\mathbb{R}} \phi(u) d\mu \right| \leq \frac{2^{Q+N+1}}{NR^N} \left\{ \max_{|x| \geq R/2} |\phi(u) u^{Q+N}| + \max_{|x| \geq R/2} |\phi'(u) u^{Q+N+1}| \right\} \|\mu\|_Q \max_{x \in \mathbb{R}} |h'(x)|. \quad (4.43)$$

PROOF. Without loss of generality, we may assume that $\|\mu\|_Q = 1$. Since h is a low pass filter, $1 - h(u/R) = 0$ and $h'(u/R) = 0$ if $|u| < R/2$. If $|u| \geq R/2 \geq 2$, (4.15) implies that $|\mu|(u) \leq 2^Q |u|^Q$. Therefore, using integration by parts and (4.15), we obtain

$$\begin{aligned} \left| \int_{\mathbb{R}} \phi(u) h(u/R) d\mu(u) - \int_{\mathbb{R}} \phi(u) d\mu(u) \right| &= \left| \int_{\mathbb{R}} \left\{ \phi'(u)(1 - h(u/R)) - \frac{1}{R} \phi(u) h'(u/R) \right\} \mu(u) du \right| \\ &\leq 2^Q \int_{|u| \geq R/2} |\phi'(u) u^{Q+N+1}| \frac{du}{|u|^{N+1}} \\ &\quad + \max_{x \in \mathbb{R}} |h'(x)| \frac{2^Q}{R} \int_{|u| \geq R/2} |\phi(u) u^{Q+N}| \frac{du}{|u|^N}. \end{aligned} \quad (4.44)$$

We observe that since h is low pass filter, the mean value theorem implies that $\max_{x \in \mathbb{R}} |h'(x)| \geq 2$. The estimate (4.43) is now easy to deduce. \square

Lemma 4.5 *Let μ be an even measure satisfying (4.15), and $\phi : \mathbb{R} \rightarrow \mathbb{R}$ be an infinitely differentiable function supported on $[0, b^2]$ for some $b > 0$. Let $\psi : \mathbb{R} \rightarrow \mathbb{R}$ be defined by $\psi(u) = \phi(u^2)$.*

(a) *The function $\psi \in \mathcal{S}$.*

(b) *The function $\hat{\phi}$ can be extended to \mathbb{C} as an entire function of finite exponential type $\leq b^2$, and for any integer $m \geq 0$,*

$$|\hat{\phi}(z)| \leq \frac{c \exp(b^2 |\Im z|)}{|z|^m} \int_{\mathbb{R}} |\phi^{(m)}(u)| du, \quad z \in \mathbb{C} \setminus \{0\}. \quad (4.45)$$

(c) *If*

$$F(z) = \int_0^\infty \exp\left(-\frac{u^2}{4iz}\right) d\mu(u), \quad z = x + iy, \quad y < 0, \quad (4.46)$$

then

$$\hat{\mu}(\psi) = \int_{\mathbb{R}} \hat{\phi}(-x - iy) (i(x + iy)/\pi)^{-1/2} F(x + iy) dx, \quad y < 0, \quad (4.47)$$

where the principal branch of the square root is understood.

PROOF. Part (a) is easy to verify using the chain rule of differentiation several times.

Defining

$$\hat{\phi}(z) = \frac{1}{2\pi} \int_0^{b^2} \exp(-iuz) \phi(u) du, \quad z \in \mathbb{C},$$

it is clear that $\hat{\phi}$ extends the function $\hat{\phi}$ on \mathbb{R} to \mathbb{C} . The fact that this extension is an entire function of finite exponential type $\leq b^2$ and (4.45) are elementary to verify. This proves part (b).

In view of (4.45), we may apply the Fourier inversion formula to deduce that

$$\phi(u) = \exp(uy) \int_{\mathbb{R}} \hat{\phi}(-x - iy) \exp(-ixu) dx, \quad u, y \in \mathbb{R}. \quad (4.48)$$

Without loss of generality, we may assume that $\|\mu\|_Q = 1$. In this proof only, let h be an infinitely differentiable low pass filter as in Lemma 4.4 and for $u \in \mathbb{R}$, $d\mu_R(u) = h(u/R) d\mu(u)$, $R \geq 4$. Then μ_R is a function with bounded total variation on \mathbb{R} , $|\mu_R|(\mathbb{R}) \leq cR^Q$, $R \geq 2$. Hence, $\widehat{\mu_R}$ is an everywhere defined, even, and bounded function on \mathbb{R} , with $|\widehat{\mu_R}(u)| \leq cR^Q$, $u \in \mathbb{R}$. In this proof, we write

$$g_R(u) = u^{-1/2} \widehat{\mu_R}(\sqrt{u}), \quad u \in [0, \infty), \quad R \geq 4.$$

Then

$$\int_0^\infty \phi(u) g_R(u) du = 2 \int_0^\infty \phi(u^2) \widehat{\mu_R}(u) du = \int_{\mathbb{R}} \psi(u) \widehat{\mu_R}(u) du = \int_{\mathbb{R}} \hat{\psi}(u) d\mu_R(u),$$

and Lemma 4.4 shows that

$$\hat{\mu}(\psi) = \int_{\mathbb{R}} \hat{\psi}(u) d\mu(u) = \lim_{R \rightarrow \infty} \int_0^\infty \phi(u) g_R(u) du. \quad (4.49)$$

Next, we recall [59, Lemma 7.6] that

$$\exp(-x^2/2) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} \exp(-v^2/2) \exp(-ivx) dv.$$

Let $t > 0$. Letting $x = u/\sqrt{2t}$, and $v = w\sqrt{2t}$, this leads to

$$\exp(-u^2/(4t)) = \sqrt{t/\pi} \int_{\mathbb{R}} \exp(-tw^2) \exp(-i w u) dw, \quad u \in \mathbb{R}.$$

Since μ_R is a measure of bounded total variation on \mathbb{R} , we may use Fubini's theorem to conclude that

$$\int_{\mathbb{R}} \exp(-u^2/(4t)) d\mu_R(u) = \sqrt{t/\pi} \int_{\mathbb{R}} \exp(-w^2 t) \widehat{\mu_R}(w) dw = \sqrt{t/\pi} \int_0^\infty \exp(-tv) g_R(v) dv, \quad (4.50)$$

and hence, by analytic continuation that

$$F_R(z) := \int_{\mathbb{R}} \exp(-u^2/(4iz)) d\mu_R(u) = \sqrt{iz/\pi} \int_0^\infty \exp(-izv) g_R(v) dv, \quad z \in \mathbb{C}, \Im z < 0. \quad (4.51)$$

In view of (4.45), we see that $\exp(uy)\hat{\phi}(-x-iy)g_R(u)$ is absolutely integrable with respect to $dxdu$ on $\mathbb{R} \times [0, \infty)$ if $y < 0$. Using (4.48), (4.51), we deduce that for $y < 0$, $R \geq 2$,

$$\begin{aligned} \int_0^\infty \phi(u) g_R(u) du &= \int_0^\infty g_R(u) \left\{ \exp(uy) \int_{\mathbb{R}} \hat{\phi}(-x-iy) \exp(-ixu) dx \right\} du \\ &= \int_{\mathbb{R}} \hat{\phi}(-x-iy) \left\{ \int_0^\infty \exp(-i(x+iy)u) g_R(u) du \right\} dx \\ &= \int_{\mathbb{R}} \hat{\phi}(-x-iy) (i(x+iy)/\pi)^{-1/2} F_R(x+iy) dx. \end{aligned} \quad (4.52)$$

If $z = x + iy$, $y < 0$, then it is elementary calculus to check that for every $L > 0$, $|u^L \exp(-u^2/(4iz))| \leq c(L)|z|^L|y|^{-L/2}$, and hence, (4.45) implies that for such z ,

$$|\hat{\phi}(-x-iy)(i(x+iy)/\pi)^{-1/2} F_R(x+iy)| \leq c(L, y)|z|^{-2}.$$

Therefore, using Lemma 4.4, the dominated convergence theorem, and (4.49), we arrive at (4.47) by letting $R \rightarrow \infty$ in (4.52). This proves part (c). \square

Lemma 4.6 *Let μ be an even measure satisfying (4.15) and (4.16), and F be defined by (4.46). Then*

$$|F(x+iy)| \leq c|z|^A \exp(4r^2) \exp(4r^2 y), \quad y \leq -1, \ x \in \mathbb{R}, \quad (4.53)$$

where A is the least integer $\geq \max(C, Q)$.

The proof of this lemma requires the Phragmén–Lindelöf theorem [49, Vol. II, Theorem 7.5], the required part of which we reproduce for the convenience of the reader.

Theorem 4.6 *Let D be the interior of an angle of $\alpha\pi$, $0 \leq \alpha \leq 2$, with boundary Γ , and f be analytic on D . With*

$$M(f, v) = \sup_{|z|=v, z \in D} |f(z)|,$$

let

$$\limsup_{z \rightarrow \xi} |f(z)| \leq C < \infty, \quad \xi \in \Gamma \cap \mathbb{C}, \quad (4.54)$$

and

$$\liminf_{v \rightarrow \infty, v \in D} \frac{\log \log M(f, v)}{\log v} < 1/\alpha, \quad (4.55)$$

then

$$|f(z)| \leq C, \quad z \in D. \quad (4.56)$$

PROOF OF LEMMA 4.6. We will use the Phragmén–Lindelöf theorem with

$$G(z) = \exp(4r^2 iz) z^{-A} F(z), \quad z \in \mathbb{C}, \Im z < 0,$$

in place of f . We note that G is analytic in the domain as indicated. The estimate (4.16) implies that

$$|F(-it)| = \left| \int_0^\infty \exp(-u^2/(4t)) d\mu(u) \right| \leq ct^C \exp(-4r^2 i(-it)), \quad t \geq 1/4.$$

In particular,

$$|G(z)| \leq c, \quad z = -it, \quad t \geq 1/4. \quad (4.57)$$

Also, using (4.29) and (4.15), we deduce that for $z = x - iy$, $x \in \mathbb{R}$, $y > 0$,

$$\begin{aligned} |F(x - iy)| &\leq \int_0^\infty \left| \exp\left(-\frac{u^2}{4i(x - iy)}\right) \right| d|\mu|(u) = \int_0^\infty \exp(-u^2 y/(4|z|^2)) d|\mu|(u) \\ &= \frac{y}{2|z|^2} \int_0^\infty u \exp(-u^2 y/(4|z|^2)) |\mu|(u) du \\ &\leq \frac{cy}{|z|^2} \int_0^\infty u^{Q+1} \exp(-u^2 y/(4|z|^2)) du \\ &= \frac{c|z|^Q}{y^{Q/2}} \int_0^\infty u^{Q/2} e^{-u} du. \end{aligned} \quad (4.58)$$

Using this estimate with $z = x - i$, we obtain that

$$|G(z)| = |\exp(4r^2 iz) z^{-A} F(z)| \leq c \exp(4r^2), \quad z = x - i, \quad x \in \mathbb{R}. \quad (4.59)$$

The estimate (4.58) shows further that the condition (4.55) is satisfied if $z \in \mathbb{C}$, $\Im z \leq -1$. Thus, we may apply the Phragmén–Lindelöf theorem with G , once taking D to be the sector bounded by $\{z = x - i, x \geq 0\}$ and $\{z = -it, t \geq 1\}$ and then with the sector bounded by $\{z = x - i, x \leq 0\}$ and $\{z = -it, t \geq 1\}$ to deduce (4.53) from (4.57) and (4.59). \square

PROOF OF THEOREM 4.5. Let $0 < b < 2r$, $\epsilon > 0$ and $\psi \in \mathcal{S}$ be an even function supported on $[-b, -\epsilon] \cup [\epsilon, b]$. Then there is an infinitely differentiable function ϕ supported on $[\epsilon^2, b^2]$ such that $\psi(u) = \phi(u^2)$. With F as in (4.46), Lemma 4.5 and Lemma 4.6 imply that for $y < -1$,

$$|\hat{\mu}(\psi)| = \left| \int_{\mathbb{R}} \hat{\phi}(-x - iy) (i(x + iy)/\pi)^{-1/2} F(x + iy) dx \right| \leq c \exp(4r^2) \exp(-(4r^2 - b^2)|y|).$$

Since $0 < b < 2r$, the right hand side of this inequality tends to 0 as $y \rightarrow -\infty$. Hence $\hat{\mu}(\psi) = 0$. Since ϵ , b , and ψ were arbitrary, this implies that $\hat{\mu}$ is supported on $[-\infty, -2r] \cup [2r, \infty)$. \square

Finally, we are in a position to put together the results in the three steps above to complete the proof of Theorem 4.3.

PROOF OF THEOREM 4.3. In view of Theorem 4.5, the condition (4.16) implies that the support of $\hat{\mu}$ is contained in $(-\infty, -2r] \cup [2r, \infty)$. Theorem 4.3 follows immediately from Theorem 4.4. \square

References

- [1] R. Askey. *Orthogonal polynomials and special functions*. SIAM, 1975.
- [2] M. Belkin, I. Matveeva, and P. Niyogi. Regularization and semi-supervised learning on large graphs. In *Learning theory*, pages 624–638. Springer, 2004.
- [3] M. Belkin and P. Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural computation*, 15(6):1373–1396, 2003.
- [4] M. Belkin and P. Niyogi. Semi-supervised learning on Riemannian manifolds. *Machine learning*, 56(1-3):209–239, 2004.
- [5] M. Belkin and P. Niyogi. Towards a theoretical foundation for Laplacian-based manifold methods. *Journal of Computer and System Sciences*, 74(8):1289–1308, 2008.

- [6] A. L. Bertozzi and A. Flenner. Diffuse interface models on graphs for classification of high dimensional data. *Multiscale Modeling & Simulation*, 10(3):1090–1118, 2012.
- [7] S. Bochner and K. Chandrasekharan. *Fourier transforms*. Number 19 in Annals of mathematics studies. Princeton University Press, 1949.
- [8] C. K. Chui and D. L. Donoho. Special issue: Diffusion maps and wavelets. *Appl. and Comput. Harm. Anal.*, 21(1), 2006.
- [9] C. K. Chui and H. N. Mhaskar. Smooth function extension based on high dimensional unstructured data. *Mathematics of Computation*, 83(290):2865–2891, 2014.
- [10] C. K. Chui and J. Wang. Dimensionality reduction of hyperspectral imagery data for feature classification. In *Handbook of Geomathematics*, pages 1005–1047. Springer, 2010.
- [11] C. K. Chui and J. Wang. Randomized anisotropic transform for nonlinear dimensionality reduction. *GEM-International Journal on Geomathematics*, 1(1):23–50, 2010.
- [12] C. K. Chui and J. Wang. Nonlinear methods for dimensionality reduction. In *Handbook of Geomathematics*, pages 1–46. Springer, 2015.
- [13] F. Chung. Laplacians and the cheeger inequality for directed graphs. *Annals of Combinatorics*, 9(1):1–19, 2005.
- [14] R. R. Coifman and M. J. Hirn. Bi-stochastic kernels via asymmetric affinity functions. *Applied and Computational Harmonic Analysis*, 35(1):177–180, 2013.
- [15] R. R. Coifman and M. J. Hirn. Diffusion maps for changing data. *Applied and Computational Harmonic Analysis*, 36(1):79–107, 2014.
- [16] R. R. Coifman and S. Lafon. Diffusion maps. *Applied and computational harmonic analysis*, 21(1):5–30, 2006.
- [17] R. R. Coifman and M. Maggioni. Diffusion wavelets. *Applied and Computational Harmonic Analysis*, 21(1):53–94, 2006.
- [18] I. Daubechies. *Ten lectures on wavelets*, volume 61. SIAM, 1992.
- [19] L. D. David and G. Carrie. Hessian eigenmaps: new locally linear embedding techniques for high dimensional data, tr2003-08, dept. of statistics, 2003.
- [20] E. B. Davies. *Heat kernels and spectral theory*, volume 92. Cambridge University Press, 1990.
- [21] R. A. DeVore and G. G. Lorentz. *Constructive approximation*, volume 303. Springer Science & Business Media, 1993.
- [22] D. L. Donoho and C. Grimes. Image manifolds which are isometric to euclidean space. *Journal of mathematical imaging and vision*, 23(1):5–24, 2005.
- [23] D. L. Donoho, O. Levi, J.-L. Starck, and V. Martinez. Multiscale geometric analysis for 3d catalogs. In *Astronomical Telescopes and Instrumentation*, pages 101–111. International Society for Optics and Photonics, 2002.
- [24] M. Ehler, F. Filbir, and H. N. Mhaskar. Locally learning biomedical data using diffusion frames. *Journal of Computational Biology*, 19(11):1251–1264, 2012.
- [25] F. Filbir and H. N. Mhaskar. A quadrature formula for diffusion polynomials corresponding to a generalized heat kernel. *Journal of Fourier Analysis and Applications*, 16(5):629–657, 2010.
- [26] F. Filbir and H. N. Mhaskar. Marcinkiewicz-Zygmund measures on manifolds. *Journal of Complexity*, 27(6):568–596, 2011.
- [27] F. Filbir, R. Hielscher, and W.R. Madych. Reconstruction from circular and spherical mean data. *Applied and Computational Harmonic Analysis*, 29(1):111–120, 2010.

- [28] J. Friedman and J.-P. Tillich. Wave equations for graphs and the edge-based laplacian. *Pacific Journal of Mathematics*, 216(2):229–266, 2004.
- [29] G. A. Gidelew. *Topics in harmonic analysis on combinatorial graphs*. PhD thesis, Drexel University, 2014.
- [30] F. Girosi and T. Poggio. Networks and the best approximation property. *Biological cybernetics*, 63(3):169–176, 1990.
- [31] I. Gohberg and M. G. Krein. *Introduction to the theory of linear nonselfadjoint operators*, volume 18. American Mathematical Soc., 1969.
- [32] A. Grigoryan. Upper bounds of derivatives of the heat kernel on an arbitrary complete manifold. *Journal of Functional Analysis*, 127(2):363–389, 1995.
- [33] A. Grigor’yan. Gaussian upper bounds for the heat kernel on arbitrary manifolds. *history*, 4(2exp):4t, 1997.
- [34] A. Grigor’yan. Estimates of heat kernels on Riemannian manifolds. *London Math. Soc. Lecture Note Ser*, 273:140–225, 1999.
- [35] A. Grigor’yan, J. Hu, and K.-S. Lau. Heat kernels on metric measure spaces. In *Geometry and Analysis of Fractals*, pages 147–207. Springer, 2014.
- [36] P. R. Halmos and V. S. Sunder. *Bounded integral operators on L^2 spaces*, volume 96. Springer Science & Business Media, 2012.
- [37] D. K. Hammond, P. Vandergheynst, and R. Gribonval. Wavelets on graphs via spectral graph theory. *Applied and Computational Harmonic Analysis*, 30(2):129–150, 2011.
- [38] X. He, S. Yan, Y. Hu, P. Niyogi, and H.-J. Zhang. Face recognition using Laplacianfaces. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 27(3):328–340, 2005.
- [39] R. A. Horn and C. R. Johnson. *Matrix analysis*. Cambridge university press, 2012.
- [40] P. W. Jones, M. Maggioni, and R. Schul. Manifold parametrizations by eigenfunctions of the Laplacian and heat kernels. *Proceedings of the National Academy of Sciences*, 105(6):1803–1808, 2008.
- [41] P. W. Jones, M. Maggioni, and R. Schul. Universal local parametrizations via heat kernels and eigenfunctions of the Laplacian. *Ann. Acad. Sci. Fenn. Math.*, 35:131–174, 2010.
- [42] W. H. Kim, V. Singh, M. K. Chung, C. Hinrichs, D. Pachauri, O. C. Okonkwo, S. C. Johnson, Alzheimer’s Disease Neuroimaging Initiative, et al. Multi-resolutional shape features via non-euclidean wavelets: Applications to statistical analysis of cortical thickness. *NeuroImage*, 93:107–123, 2014.
- [43] Y. A. Kordyukov. L^p -theory of elliptic differential operators on manifolds of bounded geometry. *Acta Applicandae Mathematica*, 23(3):223–260, 1991.
- [44] J. Korevaar. *Tauberian theory: a century of developments*, volume 329. Springer Science & Business Media, 2004.
- [45] S. S. Lafon. *Diffusion maps and geometric harmonics*. PhD thesis, Yale University, 2004.
- [46] Z. Li, U. Park, and A. K. Jain. A discriminative model for age invariant face recognition. *Information Forensics and Security, IEEE Transactions on*, 6(3):1028–1037, 2011.
- [47] Z.-M. Ma and M. Röckner. *Introduction to the theory of (non-symmetric) Dirichlet forms*. Springer, 1992.
- [48] M. Maggioni and H. N. Mhaskar. Diffusion polynomial frames on metric measure spaces. *Applied and Computational Harmonic Analysis*, 24(3):329–353, 2008.
- [49] A. I. Markushevich. *Theory of functions of a complex variable*, volume 296. American Mathematical Soc., 2005.
- [50] H. N. Mhaskar. Eignets for function approximation on manifolds. *Applied and Computational Harmonic Analysis*, 29(1):63–87, 2010.

- [51] H. N. Mhaskar. A generalized diffusion frame for parsimonious representation of functions on data defined manifolds. *Neural Networks*, 24(4):345–359, 2011.
- [52] H. N. Mhaskar and J. Prestin. Polynomial frames: a fast tour. *Approximation Theory XI: Gatlinburg*, pages 101–132, 2004.
- [53] T. B. Moeslund. *Introduction to video and image processing: Building real systems and applications*. Springer Science & Business Media, 2012.
- [54] S. Mousazadeh and I. Cohen. Embedding and function extension on directed graph. *Signal Processing*, 111:137–149, 2015.
- [55] Y. Nakatsukasa, Z. Bai, and F. Gygi. Optimizing Halley’s iteration for computing the matrix polar decomposition. *SIAM Journal on Matrix Analysis and Applications*, 31(5):2700–2720, 2010.
- [56] S. M. Nikolskii. *Approximation of functions of several variables and imbedding theorems*. Springer Verlag, 1975.
- [57] S. T. Roweis and L. K. Saul. Nonlinear dimensionality reduction by locally linear embedding. *Science*, 290(5500):2323–2326, 2000.
- [58] W. Rudin. *Real and complex analysis*. Tata McGraw-Hill Education, 1987.
- [59] W. Rudin. Functional analysis. international series in pure and applied mathematics, 1991.
- [60] A. Sikora. Riesz transform, Gaussian bounds and the method of wave equation. *Mathematische Zeitschrift*, 247(3):643–662, 2004.
- [61] A. Singer. From graph to manifold Laplacian: The convergence rate. *Applied and Computational Harmonic Analysis*, 21(1):128–134, 2006.
- [62] G Szegő. Orthogonal polynomials. In *Colloquium publications/American mathematical society*, volume 23. Providence, 1975.
- [63] J. B. Tenenbaum, V. De Silva, and J. C. Langford. A global geometric framework for nonlinear dimensionality reduction. *Science*, 290(5500):2319–2323, 2000.
- [64] Y. van Gennip, B. Hunter, R. Ahn, P. Elliott, K. Luh, M. Halvorson, S. Reid, M. Valasik, J. Wo, G. E. Tita, et al. Community detection using spectral clustering on sparse geosocial data. *SIAM Journal on Applied Mathematics*, 73(1):67–83, 2013.
- [65] G. N. Watson. *A treatise on the theory of Bessel functions*. Cambridge university press, 1995.
- [66] K. Q. Weinberger, B. D. Packer, and L. K. Saul. Nonlinear dimensionality reduction by semidefinite programming and kernel matrix factorization. In *Proceedings of the tenth international workshop on artificial intelligence and statistics*, pages 381–388. Citeseer, 2005.
- [67] Z.-Y. Zhang and H.-Y. Zha. Principal manifolds and nonlinear dimensionality reduction via tangent space alignment. *Journal of Shanghai University (English Edition)*, 8(4):406–424, 2004.